Albert Newen[1] and
Carlos Montemayor[2]

The ALARM Theory of Consciousness

A Two-Level Theory of Phenomenal Consciousness

Abstract: *The scientific investigation of consciousness generates new findings at a rapid pace. We argue that we need a novel theoretical framework, which we call the ALARM theory of consciousness, in order to account for all central observations. According to this theory, we need to distinguish two levels of consciousness, namely basic arousal and general alertness. Basic arousal functions as a specific alarm system, keeping a biological organism alive under sudden intense threats, and general alertness enables flexible learning and behavioural strategies. This two-level theory of consciousness helps us to account for (i) recent discoveries of subcortical brain activities with a central role of thalamic processes, (ii) observations of differences in the behavioural repertoire of non-human animals indicating two types of conscious experiences. Furthermore, the framework enables us (iii) to unify the neural evidence for the relevance of subcortical processes, on the one hand, and of cortico-cortical loops, on the other, and finally (iv) to clarify the evolutionary and actual functional role of conscious experiences.*

Correspondence:
Email: cmontema@sfsu.edu

1 Institut für Philosophie II, Ruhr-Universität Bochum, Germany.
2 Philosophy Department, San Francisco State University, USA.

Journal of Consciousness Studies, **30**, No. 3–4, 2023, pp. 84–105
DOI: 10.53765/20512201.30.3.084

1. Main Deficits of Philosophical Theories of Consciousness from a Bird's Eye View

The central claim of the theory we propose (ALARM) is that there are two distinct neural regional networks that constitute consciousness at two levels. The first level is more inflexible, reflex-like, and has more ancient evolutionary origins. The second level provides flexibility, a robust connection with planning, learning, and attention routines, and plays a more fundamental role in higher cognitive functions. Both are aspects of conscious awareness — they are not separate kinds of consciousness.[3] Below we argue that, in fact, conscious awareness is always a deeply unified phenomenon in spite of this crucial distinction.

We make two initial clarifications. First, concerning the scope of our theory: our focus is phenomenal consciousness (rather than access or other kinds of consciousness) and we do not aim to account for disorders of consciousness, at least in this contribution. ALARM postulates two levels of information processing that constitute phenomenal consciousness. Second, our claims are not merely conceptual or *a priori*. We do not argue for some hypothetical or metaphysical ground that is necessary and sufficient for consciousness. Rather, we claim that the available empirical evidence shows that consciousness is constitutively determined by two levels of neural processing. This is the standard way in which other theories that postulate neural correlates of consciousness (NCC) are interpreted — as the anatomical grounds that are specifically relevant for constituting consciousness.[4]

Thus, we are aiming to exclude trivial causal background conditions (temperature, heart rate, and genetic information all causally contribute to consciousness, but no one thinks that this is of any *specific* interest to the science of conscious episodes) and we are searching for relevant causal underpinnings of conscious experience. Our specific contribution to the debate on the NCC is that an important distinction

---

3 To clarify, as mentioned in the next paragraph, we are distinguishing two levels of awareness with different roles, not two types of consciousness or two mechanisms for separate kinds of consciousness. We highlight evidence suggesting that there are distinct correlates of these aspects of phenomenal consciousness, but we are referring to the same kind of unified phenomenal consciousness. We are grateful to an anonymous reviewer for helping us to clarify this issue.

4 There are more fine-grained distinctions of NCC (Aru *et al.*, 2012) but we want to focus on the central implementations of phenomenal consciousness in a general sense.

is missing, one that has implications for behaviourally rigid and flexible aspects of consciousness which we will account for with two varieties of consciousness, namely basic arousal and general alertness.

From a bird's eye view, we can classify all theories of consciousness as either content theories or empirically-driven process theories (while some rare cases are mixed theories). Content theories argue that consciousness depends fundamentally at least on properties associated with distinct representational contents which need to be processed. In contrast, pure process theories claim that it is not a specific content but only the way of processing a content that determines whether the content is consciously experienced or not. Thus, pure process theories exclude a specific content as a constitutive component of consciousness.

We need to distinguish between two types of content theories:

1. Pure content theories: These content views endorse the claim that some kind of mental or representational state with a specific content is necessary and sufficient for contents to become conscious, e.g. qualia theories and 'primitivist' accounts.
2. Minimal content theories: These theories claim that some kind of mental or representational state with a specific content is necessary (but not sufficient) for contents to become conscious.

Pure content theories, e.g. Chalmers (1996; 2006), share the commitment that being in a state that directly represents or acquaints us with a specific content is necessary and sufficient for conscious experience. *Minimal content theories* are compatible with some specific processes being involved in producing phenomenal consciousness. But as long as they necessitate content of a specific kind being involved in constituting consciousness, that makes them content views according to our definition. For example, Jesse Prinz's (2012) AIR theory of consciousness proposes a view that involves both; namely, he argues that only some specific processes, i.e. intermediate-level *activations*, produce consciousness. But what makes these activations relevant is that they process a specific kind of content, namely intermediate-level *representations* (Marchi and Newen, 2015). Other famous theories have a similar commitment to both a kind of processing and a specific kind of content, e.g. Tye's PANIC account involves a specific content: 'Phenomenal character is one and the same as Poised, Abstract, Non-conceptual, Intentional Content' (Tye, 2000, p. 63). But this characterization of an abstract, non-conceptual, propositional content also involves a processing aspect, expressed by being poised: a state is

poised if and only if the state has the disposition to directly influence our beliefs and/or desires (*ibid.*, p. 62). The latter is not a specification of the content, but a description of a functional role of the previously specified content, i.e. it constrains a certain process. A further example of minimal content theories are the higher-order theories of consciousness (HO theories). The central claim of HO theories is that a mental state m is consciously experienced by a person if that person is in the state m and at the same time has an unconscious meta-representational state with the content *that she herself is in m*. On some versions of the theory, the unconscious meta-state is a perceptual state (Lycan, 1996). On others, it is either an occurrent thought that is realized in the relevant situation of having a subjective conscious experience (Rosenthal, 1986; 1993), or a dispositional thought (Carruthers, 2000). Although there are different aspects of processing highlighted (perceptual, occurrent, dispositional), the central common ground is that the meta-representational content is necessarily involved in producing consciousness.

It is not possible to do justice to the enormous variety of philosophical content theories here, but almost all of them suffer from the fact that they are not anchored in recent empirical observations produced during the last fifteen years. There are two core empirical observations which demand a shift from content theories to process theories of consciousness. (i) Typical first-order contents of a perceptual experience, e.g. seeing the shape and direction of a letter box, can also be processed unconsciously, as is demonstrated in the famous case of visual agnosia (Milner and Goodale, 1995); there are now many findings on a wide range of psychological phenomena supporting the conclusion that phenomenal experience cannot be explained by having specific contents because often the same normally conscious contents can sometimes be processed unconsciously, e.g. in cases of blindsight or in priming experiments (Vosgerau, Schlicht and Newen, 2008). (ii) The empirical studies of Dehaene (2014) and his colleagues demonstrate that specific processes, namely neural integration processes, seem to play a crucial role in enabling conscious experience — and the contribution of integration processes seems to be a minimal common ground, despite a variety of differences in recent theories (see below). Both observations strongly

indicate that it is not a specific kind of content,[5] but a specific type of processing that is responsible for consciousness. Therefore, process theories seem to provide the only path forward for an empirically valid theory of consciousness. This is in line with our commitment, expressed above, to aim for an empirical anchoring of our theory.

There are two prominent processing accounts, namely global neuronal workspace theory (GNWT) and integrated information theory (IIT), which are promising candidates to solve the riddle of consciousness, but we think that neither is yet a fully adequate account because they both ignore or do not appreciate adequately two important aspects of consciousness: (a) an evolutionary perspective according to which we should be able to reconstruct the basic evolutionary role of consciousness which may not be the same as the synchronic functional role everyday consciousness has nowadays, and (b) the role of neural processes in the brainstem and in thalamic areas for consciousness, confirmed by recent discoveries. We will include these dimensions into our criteria of adequacy, elaborate them, and demonstrate that we have to enrich the core idea of integration in the current versions of both GNWT and IIT.

Another argument demands a change of the conceptual framework for investigating consciousness. All theories developed so far, namely content and process theories of phenomenal consciousness which aim to explain the realization of consciousness, are presupposing a single threshold of entry or activation of consciousness making phenomenal consciousness either present or absent. We explain why this is a problematic assumption in detail below. But to quickly appreciate this point, we rely on considerations about the evolution of phenomenal consciousness. Given the complexity of the contents and brain areas involved in various types of awareness, it would be surprising that a single threshold or type of content is what matters for all kinds of phenomenal consciousness (e.g. sensorial, emotional, motivational, intellectual). We propose the ALARM theory of consciousness because it allows for a more nuanced comparison across species, based on an evaluation of the complex interaction between contents and thresholds of activation, in a more flexible and empirically

[5] If one objects that the difference between contents and processes is not clear enough since contents are realized as processes, then the relevant claim is that those processes underlying specific contents proposed by the content theories are not constitutive for consciousness (given the evidence of multiple studies concerning unconscious processing).

informed manner. We focus on two experimentally confirmed types of consciousness here.

2. The Need for a Two-Level Theory of Consciousness

We argue that almost all available theories of phenomenal consciousness, both philosophical and scientific, need some additions or modifications because they do not jointly meet the four criteria of adequacy for any theory of consciousness, which we present now. They are the result of a more precise way of formulating the principles discussed above. The first criterion of adequacy (C1: functional role) is that a thorough explanation of consciousness should include a clear *evolutionary functional role* for consciousness associated with survival as well as a *synchronic functional role* associated with the actual cognitive advantages afforded by conscious experiences. While some theories are useful in providing evolutionary explanations, none of them explicitly addresses evolutionary functions and only a few discuss specific advantages of being capable of conscious awareness at a point in time.

As an illustration, Humphrey (2011; 2022) has argued that the evolutionary purpose of consciousness is not directly tied to accurate cognition or knowledge, but rather with life enjoyment, emotional engagement, and a sense of self (2022, chapters 13 and 23). This may indeed be a very important advantage of phenomenal consciousness at the level of awareness that humans have. But we argue that other cognitive functions, more basic and evolutionarily tied with alerting the entire organism, need not be characterized in terms of a sense of self, at least not in the demanding sense that is used by Humphrey. Our approach is therefore compatible with accounts of early forms of consciousness that associate it with flexible forms of learning and cognition (Ginsburg and Jablonka, 2019), without necessarily assuming a demanding sense of self, e.g. in the form of the first-person perspective or a 'meaning of life' (Humphrey, 2011); the latter might actually depend on general alertness.[6]

---

6 See Merker, Williford and Rudrauf (2021) for an account of the first-person perspective in a more minimal sense than the one discussed by Humphrey, and in the context of criticizing IIT. See Mudrik, Faivre and Koch (2014) for evidence that many kinds of high-level information integration can occur in the absence of awareness, and Haladjian

The second criterion is that, in connection with these different functional roles from an evolutionary and synchronic perspective, we can observe two paradigmatically different ranges of behavioural abilities in the same type of situation (C2: behavioural). Thirdly, recent neuroscientific findings demonstrate that neural processes in the thalamus are fundamentally involved in conscious processing (C3: neural processing). This includes recent discoveries of *the role of thalamic processes* in a shift of conscious attention as described by Halassa and colleagues (e.g. Nakajima, Schmitt and Halassa, 2019) and the role of thalamic activations for basic conscious awareness as demonstrated by Redinbaugh *et al.* (2020). Although there is a tradition of theoretical approaches that account for thalamic processes, we think that these two recent empirical observations have not been adequately considered in their full relevance.[7] This involves the search for a satisfactory explanation of the role of thalamic processing in relation to cortical regions. These discoveries demand *a solution to the tension between the role of non-cortical and neocortical processes* for consciousness: looking at the state of the art of empirical observations it seems that we are facing paradoxical claims. On the one hand, consciousness is based on the activation of evolutionarily old brain areas (brainstem and thalamus) and can be realized independently from neocortical activations (Merker, 2007). On the other hand, consciousness is necessarily based on cortical and neocortical integration processing (Dehaene, 2014).

We can solve this tension by distinguishing two levels of phenomenal consciousness. This is in line with recent suggestions concerning

and Montemayor (2015) for the possible separate evolutionary paths of phenomenal consciousness and attention on the basis of this evidence.

7 We are not claiming that the view that the thalamus is crucial for consciousness is new. What we mean is that its early characterization by, for instance, Akert and Anderson (1951) described the thalamus as relevant for sleep–wake control. Also Crick and Koch (1992) and Llinás *et al.* (1998) discuss the thalamus, but they do it in a more general and speculative way than the contemporary models of thalamic influence. These early investigations left open the possibility that thalamic processes are only supporting conditions for the realization of consciousness. But recent investigations of thalamic processes have become a lot more detailed, such as the work by Aru *et al.* (2019), Bachmann (2021), Nakajima, Schmitt and Halassa (2019), and Redinbaugh *et al.* (2020). We also acknowledge that our account builds on this recent and important work. What we add to these models is a theoretical interpretation that shows how the processing stages described in these findings fit with our conceptual characterization of basic arousal and general alertness. And the studies we discuss in more detail in the text allow us to describe a mechanistic contribution of specific thalamic activations to the two forms of phenomenal consciousness that we propose.

the integration of non-cortical and neocortical processes (Aru *et al.*, 2019; Bachmann, Suzuki and Aru, 2020). We develop the ALARM theory since it can account for all the criteria of adequacy and dissolve the opposing empirical claims by distinguishing two levels of consciousness: *basic awareness* is realized by evolutionarily older brain areas while *general alertness* needs additional cortical and neocortical processes. We proceed by demonstrating that the two-level ALARM theory offers unique advantages with respect to behavioural, (synchronic) functional, and phenomenological perspectives as well. This includes the discussion of the fourth criterion, namely to include an adequate description of *the phenomenology of the unity of consciousness* (C4) despite its connection with rigid reactions, on the one hand, and flexible behaviour, on the other. Thus, we argue that the proposed ALARM theory of consciousness can meet all four criteria. To do this, we need to overcome the general presupposition of contemporary theories of phenomenal consciousness, namely that there is only a single threshold or a single type of processing at which either contents become conscious or neural activations produce phenomenal consciousness. We argue that we need to presuppose (at least) a dual type of processing which underlies two levels of phenomenal consciousness.

3. Key Claims and Evidence in Favour of the ALARM Theory of Consciousness

ALARM's central distinction between two levels of consciousness, namely basic awareness and general alertness, is proven to be adequate and epistemically fruitful, based on four perspectives with which we address the four criteria of adequacy: the evolutionary and synchronic functional perspective, the behavioural perspective, the neural processing perspective, and the phenomenological perspective.

3.1. The functional perspective

From the evolutionary and synchronic functional perspective (criterion C1), we can distinguish two roles of consciousness: one for evolutionarily older brain systems and an additional one for evolutionarily younger cortical structures. The evolutionarily old functional role of basic awareness is to trigger an alarm signal in the biological system which can then start immediate survival reaction programs. The evolutionarily younger (synchronic) functional role of general alertness is to enable or accelerate specific learning processes by

selecting contexts and associating contents. Both roles are in contrast to the claim of Kriegel: 'The functional role of consciousness is to give the subject just enough information to know how to quickly and effortlessly obtain rich and detailed information about her concurrent experience' (Kriegel, 2004). To highlight this functional role, he identifies consciousness with 'peripheral self-awareness' (*ibid.*). Thus, consciousness should deliver us information about our experiences. We think that this describes the role of consciousness backwards. The evolutionary functional role of consciousness is not to enable us to have a better knowledge of our mental life (inwards-directed self-awareness) but to better inform us about challenges for the body in a given environment (outwards-directed sensitivity to a cause of the experience).

Why is this plausible? To develop our argument, we discuss homeostatic processes and we presuppose that those can be adequately described as Bayesian processes. In line with Damasio (1999), we presuppose that homeostatic processes like temperature regulation are extremely relevant for survival, and they are the basis of the evolution of affective processing. Homeostatic regulation processes are realized in a kind of slow Bayesian updating which regulates, for instance, increased sweating in the case of raising temperature. One way to describe homeostatic regulation is suggested by Seth (2013): it is a dynamic adjustment of an interoceptive prior hypothesis to exteroceptive signals (temperature change) in the context of a multiplicity of sensory signals. We highlight that this mechanism is ideal for slow updates. When confronted with a radical challenge, e.g. suddenly the living being enters an area of very high temperature, the slow Bayesian updating needs to be stopped and the incoming temperature signal needs to receive 100% weight. Since the signal remains normally unconscious, a transfer into basic consciousness makes a radical difference by pitching it into the foreground and thereby stopping the slow Bayesian updating. Then the biological system can trigger an immediate survival reaction which is a big evolutionary advantage. This is the key functional role of basic awareness and therefore our theory is called the ALARM theory of consciousness.

According to ALARM, the purpose of the first conscious stage of basic awareness is to stop the slow Bayesian updating systems and start a process of full-energy focus, giving the challenging signal absolute priority, and triggering and keeping a reaction that enables or fosters survival. This theory is supported by the observation that almost all homeostatic regulation processes in the body lead to

conscious negative feelings or even pain if the automatic regulation does not keep the body in the acceptable range and survival is threatened (e.g. high fever leads to strong negative feelings, a sudden lack of oxygen hurts, if blood pressure is suddenly too high you get a headache, etc.). Basic awareness which is often realized by our pain processing system generates alarm signals about internal or external challenging states for our body: typical external challenges are standard bodily pain episodes (e.g. if someone steps bare-feet onto a thorn or touches hot metal).[8]

We argue that, on the basis of this core evolutionary functional role, conscious experiences have developed a second functional role, which is *general alertness* for stabilizing visceral conscious signals and enabling new types of learning. In line with the work of Ginsburg and Jablonka (2019), we suggest that a key ability involves unlimited associative learning. In addition, we would like to mention the ability of *one-case learning* which can be realized in a specific way as soon as it is combined with general alertness: basic awareness is already sufficient to learn simple connections, e.g. a biological system, being hurt by fire, never touches fire again. But if this is combined with general alertness, i.e. basic attention on the whole situation (combining relevant signals with successful reactions and a contextual interpretation of the signals), the system can learn and practise new responses (to extinguish the fire). In addition to basic arousal, general alertness is necessary to enable focused attention or focused cognition, which is the presupposition for new strategies of learning, e.g. unlimited associative learning and any form of explicit learning. These advanced functional roles of consciousness are described by Dehaene (2014, chapter 3) as follows: consciousness is used to keep information as long as we need it, to compress information (e.g. into non-linguistic concepts or symbols), and to transfer compressed information to new and higher levels of processing. These advanced functional roles contribute to how different dimensions of learning are enabled and this is, according to our theory, the functional role of general alertness. The problem is that Dehaene claims that these are the only functional roles of consciousness, thereby ignoring the

8 At this point, there might be a worry concerning temporal processing, namely that pain reactions, e.g. withdrawing the hand from a hot stovetop, are sometimes already triggered before conscious experience arises — as one of the reviewers helpfully observed. We account for that below.

ancient but still crucial functional role of viscerally alarming the cognitive system. ALARM distinguishes the evolutionarily old functional role of alarm-signalling and the evolutionarily younger (synchronic) functional role of enabling new dimensions of learning with the two-level theory.

Let us elaborate on this by accounting for the challenge of the sequence of temporal processing in some cases of pain reactions: if a person touches the stovetop, then there is a very fast reaction of withdrawing the hand, which is already initiated before the person is aware of the intense pain. How can this be reconciled with the ALARMing function? We think that this quick process is part of an evolutionarily older specific process which was essentially optimized when basic awareness evolved. Biological systems started out with the unconscious processing of challenging sensory stimuli and improved their survival when they associated unconsciously a fast survival reaction. It is plausible to presuppose that those associations of stimuli and survival reactions were rather specific: in the case of fire, the pain sensors in the hand were immediately triggered and strictly associated with the withdrawing reaction. If now consciousness in the form of basic arousal comes into play, then the subject not only has intense sensory stimuli activation in the very first moment the hand is on the stove but, since the conscious pain experience goes on after withdrawing the hand, the subject has a continuing signal to care for the body in relation to this injury.

Furthermore, the pain experience is a much more general indicator of an intense challenge to the body: it is activated in all cases of sudden high sensory stimulations, not only in already established contexts, i.e. it is a general indicator also for new situations. In addition, when consciousness in the form of general alertness evolved, then this enabled the improvement of the reaction patterns: now not only one and the same survival program can be activated, but new behavioural reactions can be realized to support the survival of the body. In short: non-conscious processes already enabled fast survival reactions in *specific situations*. Basic arousal established pain as a *more general indicator* of body challenges for a large variety of cases and new situations. It added the aspect of systematic care for the body for an extended time. Finally, general alertness enables multiple types of behavioural reactions to better deal with these challenges.

3.2. The behavioural perspective

On this basis, we can also distinguish two levels of behavioural abilities (C2). Humans have several hardwired behavioural programs which help us survive in cases of severe danger, e.g. flight or freeze reactions and similar visceral reactions (e.g. when being attacked by an aggressive pit bull). These behaviours radically differ from flexible and long-term behaviours like shopping, visiting a friend, or travelling. The latter are usually based on an explicitly conscious decision to do or start an activity on a certain day. Thus, awareness of the stimulus or situation is typically connected with automatic reactions: in the case of a dangerous situation, this triggers a survival behaviour of freeze and/or flight, in everyday cases it can trigger habitual behaviour like picking up a phone that is ringing. Although such reaction patterns can in principle be realized without basic awareness in a biological system, it is an important advantage if it is connected with basic awareness since, for example, intense pain experience is a general indicator of a survival threat for multiple and new situations and the basic awareness remains for some temporal extension triggering a caring for one's body even after the immediate challenge is removed. Thus, the reactive patterns, e.g. in the case of touching a hot stovetop, develop beyond the immediate automatic withdrawal of the hand into a continuing caring for the hand in addition to changing the behavioural pattern in relation to the stovetop.

Basic awareness and immediate reactions do not yet need general alertness which is necessary to enable us to actively search the environment attentively and contextually or to consciously think about alternative plans and actions. But basic awareness may also activate general alertness and thereby enable us to use our full cognitive potential, enabling flexible reactions and decisions. This includes learning by observation, i.e. learning not exclusively based on an immediate experience, but accelerated by unlimited associative learning, including operant conditioning and contextual decision-modelling. This involves both imitating a successful behaviour but also learning to avoid an unsuccessful one, and it widens the ability of learning enormously by interpreting and generalizing contents with new contextually based meanings. Learning new stable correlations enables the cognitive system to develop a wide range of behavioural responses to all situations it has to deal with. This characterization of new learning abilities connected to general alertness is inspired by the

work of Ginsburg and Jablonka (2019) that unlimited associative learning is a key ability connected to phenomenal consciousness: there are probably no exact borderlines between basic awareness and general alertness. We only insist that there are two paradigmatic levels of learning involved in basic arousal and general alertness. Basic arousal already enables an adjustment of behavioural patterns due to generalized processing of challenges with an awareness indicator, e.g. pain or fear, which has some temporal extension in a situation. With *consciousness as general alertness* we receive the ability of focused cognition and thus can develop a new range of behavioural abilities, which can be much more complex, doing justice to the specific situation, and including future planning.

3.3. The neural processing perspective

Recent neuroscientific observations (C3) require the distinction of two levels of consciousness. First, two widely discussed process accounts, GNWT and IIT, underestimate the relevance of evolutionarily older brain areas involved in awareness, including the role of the brainstem, and especially the role of thalamic processing. Concerning the role of the upper brainstem, a radically non-cortical theory of consciousness is Merker's (2007), which summarizes evidence including notable findings in hydranencephalic children showing that the brainstem can produce a basic form of consciousness without any involvement from cortical areas. This is also supported by the studies of cats with a removed cortex — they are still able to move purposefully, orient themselves to their surroundings by vision and touch (as do rodents), and are capable of solving a visual discrimination task in a T-maze (Bjursten, Norrsell and Norrsell, 1976). On the basis of this evidence, Merker argues that the NCC are essentially based on processes in the upper brainstem and thalamus.[9] We agree with this conclusion if it is understood as highlighting the processing conditions for basic awareness. But since we also take the evidence in favour of GNWT seriously, it follows that consciousness is also systematically dependent on cortical and neocortical activations. We argue that the latter activations concern general alertness. Both claims are supported by recent empirical studies.

---

[9] Along the same lines, highlighting the central role of brainstem, there is the recent work of Solms (2021).

In fact, recent evidence shows that there are two types of thalamic processes involved in triggering the two levels of consciousness characterized above. For *basic awareness*, findings by Redinbaugh *et al.* (2020) confirm that thalamic activations are necessary for conscious awareness. In particular, deep brain stimulation of the central lateral thalamus (CL) in anaesthetized macaques had the effect of waking them up (*ibid.*). This stimulation functions as an on-off switch. Stopping the stimulus had the immediate consequence for the apes of falling back to sleep or into anaesthesia again. Furthermore, the state they were in with stimulation can best be described as a state of basic awareness, i.e. they react to perceptual stimuli but they were not alert enough to be able to do any more complex cognitive tasks that they could do in normal wakefulness. This finding strongly suggests that the CL is constitutive for consciousness in the form of basic awareness. Furthermore, in comparing anaesthetic states with conscious states this study highlights the relevance of at least minimal thalamo-cortical loops as an essential part of the neural correlates of conscious experience: 'our study provides empirical evidence for a circuit-level mechanism of consciousness with special emphasis on the reciprocal interaction between CL and deep cortical layers, which may serve as a common target of anesthetic drugs' (*ibid.*).

Regarding *general alertness*, we know from the impressive work of Dehaene (2014) that there is abundant evidence confirming that standard everyday consciousness with focused attention in humans depends on extended cortico-cortical interactions, described as integration processes within a global workspace. We suggest that these 'large-scale' cortico-cortical interactions are especially relevant for the advanced type of consciousness which we characterize as general alertness.

How are basic arousal and general alertness interconnected? We suggest that thalamic processes play a gateway role and that they do so in a *twofold manner*. Evidence for the relevance of thalamic processes in the interaction between basic awareness and general alertness can be found in Stehberg *et al.* (2001). This study identifies shared circuits for visceral signal processing and attention in the thalamus. Furthermore, the CL triggers basic awareness (as confirmed by Redinbaugh *et al.*, 2020), serving as the foundation for minimal thalamo-cortical interactions. A *vertical processing loop* combines brainstem activations via the regulatory role of the CL with some minimal thalamo-cortical activations. Basic awareness is boosted into general alertness through extensive cortico-cortical and thalamo-

cortical processes, which can be described as *horizontal processing loops* dependent on a different part of thalamus: the thalamic reticular nucleus (TRN). The idea of an interaction of vertical processes connecting thalamic processes with cortical areas, on the one hand, and horizontal processes realizing interactions across the whole cortex and neocortex, from a neural perspective, is already present in Aru *et al.* (2019) and Bachmann, Suzuki and Aru (2020). More details about our account and how it differs from their view are discussed below.

Remarkable studies by Halassa and colleagues (Nakajima, Schmitt and Halassa, 2019; Wells *et al.*, 2016; Wimmer *et al.*, 2015) demonstrate the central role of the TRN in the regulation of elaborate and sustained conscious attention. Rats had to focus either on acoustic or visual stimuli presented at the same time and respond with a trained behaviour to receive a reward. Selective attention to visual or auditory stimuli in order to switch task-salient responses requires sustained general alertness. The underlying neural processes of this kind of learning in mice involve bidirectional interactions between the TRN and the neocortex. Thalamic processes activate the neocortical processes, especially prefrontal cortex, which then steer the TRN to select the relevant sensory modality (top-down activations from the cortex to the thalamus). Thus, a bidirectional thalamo-cortical loop enables general alertness.

In sum, there is a basic vertical loop of neural processes connecting brainstem activation via the thalamus with other areas in the limbic system and minimal cortical activations. This is the core mechanism underlying basic awareness, enabling the alarm system. In the case of general alertness, we have the following situation: triggered by this vertical loop, there are in addition various horizontal loops of neural processes which involve cortico-cortical as well as thalamo-cortical interactions. The combination of both vertical and horizontal processing loops enables general alertness, thereby making possible new methods of learning, such as explicit learning, generalization, and abstraction. As a consequence, this strongly enriches the range of contents of consciousness, but it would be a mistake to infer from this that these contents are constitutive of consciousness. They can only function as indicators of consciousness, e.g. for the question when to attribute consciousness to non-linguistic or non-human animals.

Other recent findings further confirm ALARM's insights. As already mentioned, ALARM's conceptual framework fits nicely with recent observations summarized by the so-called *dendritic integration theory* (Aru *et al.*, 2019; Aru, Suzuki and Larkum, 2020; Bachmann,

Suzuki and Aru, 2020). According to it, consciousness in mammals depends on large-range cortical layer 5 pyramidal cells. If their thalamic component is blocked, then the activities in the lower part of the pyramidal cell still enable the processing of stimuli, but these contents remain unconscious and without context specificity (Aru, Suzuki and Larkum, 2020). If thalamic activities enable a vertical loop of neural interaction between the lower and the upper part of a single pyramidal cell, then the previously unconscious signal is boosted into awareness. This confirms the key role of thalamic processes as a gateway to enable consciousness. Our account nicely overlaps with the claim of an integration of vertical thalamo-cortical interactions and horizontal cortico-cortical interactions. But our account differs in that we want to explicitly integrate more general or larger vertical processing loops, including early brainstem activations (relevant for homeostatic processes) — especially accounting for the observations described by Merker (2007) and Solms (2021) — which via thalamic activations are connected to cortical areas, and we do not presuppose that this needs to be modulated by cortical layer 5 pyramidal cells as the only realizing base, even if those play an important role in producing human consciousness. We allow for a larger variety of the realization base to integrate consciousness in non-human animals with brain organizations which are different from the human brain, e.g. in birds and fish; but this cannot be explained in detail here. We only want to highlight one recent investigation of a variant of neural integration processes discovered already in mammals: it is shown that an integration of the thalamus and striatum with the parietal cortex is a realization base of consciousness in macaques (Afrasiabi *et al.*, 2021). Thus, already in mammals we observe a realization base of consciousness without the involvement of frontal cortex areas, while those are claimed to be necessary according to several recent theories of consciousness, especially GNWT.

We can now add the characterization of the two distinct *functional roles* (C1) of the two levels of processing. Basic awareness is based on processes (i) which still have a rather high threshold for stimuli to be processed, i.e. only very salient stimuli are processed. (ii) The integration processes remain anatomically constrained (local) and mainly connect brainstem areas with the thalamus and the limbic system. (iii) The activated response behaviour is a rather rigid behavioural program, typically for survival. In contrast, general alertness is based on processes (i) which have a rather low threshold for stimuli to be processed, i.e. various stimuli are processed contextually.

(ii) The integration processes involve widespread (global and horizontal) cortical activation, enabling rich informational associations. (iii) The behavioural response is quite flexible and can be adjusted in different situations, based on complex learning processes. Thus, this functional characterization strongly supports a two-level account of consciousness which we offer with the ALARM theory, rather than a single threshold for conscious awareness.

3.4. The phenomenology and unity of consciousness

ALARM also fits well with observations from a phenomenological perspective (C4). Experiencing an itch, a fever, or an unspecific pain is experienced quite differently from our experience when listening to a piece of music or thinking about plans for next week. Our sensory stimuli are experienced at two levels which are also distinguishable by different connections to the motor system: sensory stimuli can be rather unfocused like an itch and trigger an automatic behaviour like scratching. The unfocused processing is closely connected with a behavioural program. Or sensory stimuli can lead to rather focused cognition like an attentive perception of a new laptop which is less closely connected to the motor system. Conscious perception and even more conscious thinking is partially decoupled from the motor system (or connected to a multiplicity of affordances) and opens the floor for new types of response behaviour. Basic awareness produces experiences with a phenomenal urge to act while those based on general alertness are much more decoupled from such an urge, thereby enabling flexible behaviour.

It could be objected that ALARM proposes a disjunctive theory of consciousness that is incompatible with the unity of consciousness (Tye, 2003), supported by the phenomenology and nature of subjective conscious awareness. There are two responses that suffice to address this objection. First, ALARM claims that there are two levels of phenomenal consciousness, one more primitive than the other, on the basis of the latest empirical evidence. This claim is compatible with the unity of consciousness because ALARM is not postulating two different types of consciousness, as for example Block (1995) does with his influential distinction between access and phenomenal consciousness. ALARM gives an account of a single type, namely phenomenal, consciousness and it distinguishes two levels of phenomenal experience (basic arousal and general alertness) which are implemented at two stages of processing: a thalamic, non-cortical

level, and a cortical level. In fact, we believe that ALARM offers a more nuanced account of the unity of consciousness because it shows how more visceral and vivid experiences can be integrated with a more stable and general type of alertness, both within the same field of phenomenal consciousness. With respect to content, an interesting possibility is that the contents of basic awareness are more dependent on immediate recognitional capacities and reactions than the contents of general alertness, associated with flexible and stable forms of inferential and conceptual reasoning. On our account, both contents can coexist in a unified field of phenomenal consciousness.

Second, the claim that the unity of consciousness entails a single threshold or level of conscious processing is implausible, even on phenomenological grounds. But for it to be made as a principled objection to ALARM, it would need to be shown that the empirical evidence is compatible with this monolithic understanding of consciousness. We propose that the empirical evidence strongly disconfirms such a monolithic approach (see Bachmann, 2000). Thus, instead of relying on a single threshold of conscious activation, ALARM postulates two levels of processing.

4. Main Objections

One may object that the notion of 'alarm' behind the present theory is too vague to be scientifically or theoretically useful. After all, many different and contrasting systems can be understood as alarms to the system, many of which can occur unconsciously (e.g. signals concerning danger that are processed unconsciously). Moreover, an alarm signal in the system does not seem to bring a unique evolutionary advantage because the problem is that even basic life forms such as bacteria may be understood as having 'alarm systems' for their survival. In response, ALARM does not include that all signals regarding danger or survival are essentially related to phenomenal consciousness. On the contrary, only signals that are processed through the thalamo-cortical network and which are susceptible to being integrated in the two-level system described above count as essentially related to phenomenal consciousness. This is both theoretically and scientifically useful. It is theoretically useful because, as mentioned, ALARM brings more nuance to the debate about the nature of phenomenal consciousness than content and standard

process views, both of which are single threshold theories.[10] And finally, ALARM is scientifically useful because it is one of the few theories that explicitly appeals to all the relevant findings in neuroscience by integrating evidence from thalamic and cortical activations.

It could also be objected that it is learning from alarm signals, rather than alarm systems themselves, that is constitutive of phenomenal consciousness. In particular, phenomenal consciousness seems to be associated with a set of capacities that allow for very flexible and general types of learning which may be characterized as unlimited associative learning (Ginsburg and Jablonka, 2019). While ALARM is compatible with these findings about learning, this objection misses the point that the purpose of phenomenal consciousness cannot be simply learning flexibly, but, crucially, that this learning needs to be related to arousal and older systems concerning homeostatic processing, as indicated by the thalamic evidence.

5. Conclusion

The two-level ALARM theory is a promising account because it meets the challenges that a scientifically informed theory of consciousness must meet. As mentioned, these challenges are that such a theory: (1) should be able to describe a plausible functional role of consciousness from an evolutionary and from a synchronic perspective; (2) must incorporate the findings concerning the role of thalamic processes; (3) should solve the tension between non-cortical and cortical processes; and (4) should account for the unity of consciousness, and explain the distinction between rigid reactions and flexible behaviour from a phenomenological perspective.

We argued that the evolutionarily basic functional role of consciousness is to trigger a state of alarm by stopping slow Bayesian updating and give a challenging signal 100% weight which then triggers a survival program or habituated behavioural patterns. And we distinguish this basic arousal from general alertness as two levels of phenomenal consciousness. ALARM can nicely account for the thalamic processes: activations of the central lateral thalamus (CL) are

[10] A notable exception is Bachmann's (2000) microgenetic account, according to which content does not become conscious instantaneously or all at once, but instead there is a gradual process. Our approach is sympathetic to Bachmann's account, but our emphasis is on the lack of theoretical work in the extant literature on the two levels of phenomenal consciousness under discussion.

the modulating factor that enables basic arousal. Basic arousal is boosted into general alertness through extensive cortico-cortical and thalamo-cortical processes, which can be described as horizontal processing loops that are modulated by a different part of thalamus: the thalamic reticular nucleus (TRN). Within our theory we can nicely describe the role of vertical neural processes combining non-cortical areas like the brainstem with cortical areas in their role for consciousness. ALARM is compatible with and benefits from the insights of the dendritic integration theory (Aru, Suzuki and Larkum, 2020): the latter highlights the modulating role of single pyramidal cells. The thalamic part of these extended cells combines brainstem areas with thalamic and cortical areas and thus enables us to understand at least how, in mammals with pyramidal cells, the non-cortical and the cortical processes are systematically connected. And this interaction seems to be important to boost basic arousal into general alertness at least in the case of humans. We remain explicitly open to having a more general realization base for vertical and horizontal neural processing loops in non-human animals.

From an empirical point of view, the compatibility of ALARM with the recent dendritic integration theory opens the door to test ALARM as a theoretical framework in humans and to investigate unexplored details. In particular, ALARM characterizes theoretically processes concerning conscious integration and attention modulation that can be tested at the neural level, and on which the dendritic integration theory is silent. Thus, we offer a theoretical framework that fits with recent empirical discoveries and which can be used for systematic empirical predictions and investigations.

Acknowledgments

The authors contributed equally to the content of this paper. They would like to thank two anonymous reviewers and the audiences at the Science of Consciousness Conference in Interlaken, 2019, and the Cognitive Science Society meeting in 2021, particularly to Colin Allen, Tecumseh Fitch, and Eva Jablonka.

References

Afrasiabi, M., Redinbaugh, M.J., Phillips, J.M., Kambi, N.A., Mohanta, S., Raz, A., Haun, A.M. & Saalmann, Y.B. (2021) Consciousness depends on integration between parietal cortex, striatum, and thalamus, *Cell Systems*, **12** (4), pp. 363–373. doi: 10.1016/j.cels.2021.02.003

Akert, K. & Anderson, B. (1951) Experimenteller beitrag zur physiologie des nucleus caudatus, *Acta Physiologica Scandinavica*, **22**, pp. 281–298.

Aru, J., Bachmann, T., Singer, W. & Melloni, L. (2012) Distilling the neural correlates of consciousness, *Neuroscience & Biobehavioral Reviews*, **36** (2), pp. 737–746. doi: 10.1016/j.neubiorev.2011.12.003

Aru, J., Suzuki, M., Rutiku, R., Larkum, M.E. & Bachmann, T. (2019) Coupling the state and contents of consciousness, *Frontiers in Systems Neuroscience*, **30** (13), art. 43. doi: 10.3389/fnsys.2019.00043

Aru, J., Suzuki, M. & Larkum, M.E. (2020) Cellular mechanisms of conscious processing, *Trends in Cognitive Sciences*, **24** (10), pp. 814–825. doi: 10.1016/j.tics.2020.07.006

Bachmann, T. (2000) *Microgenetic Approach to the Conscious Mind*, Amsterdam: John Benjamins.

Bachmann, T. (2021) Representational 'touch' and modulatory 'retouch' — two necessary neurobiological processes in thalamocortical interaction for conscious experience, *Neuroscience of Consciousness*, **2**, niab045. doi: 10.1093/nc/niab045

Bachmann, T., Suzuki, M. & Aru, J. (2020) Dendritic integration theory: A thalamocortical theory of state and content of consciousness, *Philosophy and the Mind Sciences*, **1** (II), art. 2.

Bjursten, L.-M., Norrsell, K. & Norrsell, U. (1976) Behavioural repertory of cats without cerebral cortex from infancy, *Experimental Brain Research*, **25** (2). doi: 10.1007/BF00234897

Block, N. (1995) On a confusion about a function of consciousness, *Behavioral and Brain Sciences*, **18** (2), pp. 227–247. doi: 10.1017/S0140525X00038188

Carruthers, P. (2000) *Phenomenal Consciousness: A Naturalistic Theory*, 1st ed., Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511487491

Chalmers, D.J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, New York: Oxford University Press.

Chalmers, D.J. (2006) The foundations of two-dimensional semantics, in Garcia-Carpintero, M. & Macia, J. (eds.) *Two-Dimensional Semantics: Foundations and Applications*, pp. 55–140, Oxford: Oxford University Press.

Crick, F. & Koch, C. (1992) The problem of consciousness, *Scientific American*, **267**, pp. 152–159. doi: 10.1038/scientificamerican0992-152

Damasio, A.R. (1999) *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, 1st ed., San Diego, CA: Harcourt Brace.

Dehaene, S. (2014) *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*, New York: Penguin.

Ginsburg, S. & Jablonka, E. (2019) *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*, Cambridge, MA: MIT Press.

Haladjian, H.H. & Montemayor, C. (2015) On the evolution of conscious attention, *Psychonomic Bulletin & Review*, **22** (3), pp. 595–613.

Humphrey, N. (2011) *Soul Dust: The Magic of Consciousness*, Princeton, NJ: Princeton University Press.

Humphrey, N. (2022) *Sentience: The Invention of Consciousness*, Oxford: Oxford University Press.

Kriegel, U. (2004) The functional role of consciousness: A phenomenological approach, *Phenomenology and the Cognitive Sciences*, **3** (2), pp. 171–193. doi: 10.1023/B:PHEN.0000040833.23356.6a

Llinás, R., Ribary, U., Contreras, D. & Pedroarena, C. (1998) The neuronal basis for consciousness, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **353** (1377), pp. 1841–1849. doi: 10.1098/rstb.1998.0336

Lycan, W. (1996) *Consciousness and Experience*, Cambridge, MA: MIT Press.

Marchi, F. & Newen, A. (2015) The cognitive foundations of visual consciousness: Why should we favor a processing approach?, *Phenomenology and the Cognitive Sciences*, **15** (2), pp. 247–264 doi: 10.1007/s11097-015-9425-z

Merker, B. (2007) Consciousness without a cerebral cortex: A challenge for neuroscience and medicine, *The Behavioral and Brain Sciences*, **30** (1), pp. 63–81; discussion 81–134. doi: 10.1017/S0140525X07000891

Merker, B., Williford, K. & Rudrauf, D. (2021) The integrated information theory of consciousness: A case of mistaken identity, *Behavioral and Brain Sciences*, **45**, pp. 1–72. doi: 10.1017/S0140525X21000881

Milner, A.D. & Goodale, M.A. (1995) *The Visual Brain in Action*, Oxford: Oxford University Press.

Mudrik, L., Faivre, N. & Koch, C. (2014) Information integration without awareness, *Trends in Cognitive Sciences*, **18** (9), pp. 488–496. doi: 10.1016/j.tics.2014.04.009

Nakajima, M., Schmitt, L.I. & Halassa, M.M. (2019) Prefrontal cortex regulates sensory filtering through a basal ganglia-to-thalamus pathway, *Neuron*, **103** (3), pp. 445–458. doi: 10.1016/j.neuron.2019.05.026

Prinz, J. (2012) *The Conscious Brain*, New York: Oxford University Press.

Redinbaugh, M.J., Phillips, J.M., Kambi, N.A., Mohanta, S., Andryk, S., Dooley, G.L., Afrasiabi, M., Raz, A. & Saalmann, Y.B. (2020) Thalamus modulates consciousness via layer-specific control of cortex, *Neuron*, **106** (1), pp. 66–75. doi: 10.1016/j.neuron.2020.01.005

Rosenthal, D. (1986) Two concepts of consciousness, *Philosophical Studies*, **49**, pp. 329–359.

Rosenthal, D. (1993) Thinking that one thinks, in Davies, M. & Humphreys, G. eds. *Consciousness: Psychological and Philosophical Essays*, Oxford: Blackwell.

Seth, A.K. (2013) Interoceptive inference, emotion, and the embodied self, *Trends in Cognitive Science*, **17** (11), pp. 565–573. doi: 10.1016/j.tics.2013.09.007

Solms, M. (2021) *The Hidden Spring: A Journey to the Source of Consciousness*, 1st ed., New York: W.W. Norton & Company.

Stehberg, J., Acuña-Goycolea, C., Ceric, F. & Torrealba, F. (2001) The visceral sector of the thalamic reticular nucleus in the rat, *Neuroscience*, **106** (4), pp. 745–755. doi: 10.1016/s0306-4522(01)00316-5

Tye, M. (2000) *Consciousness, Color, and Content*, Cambridge, MA: MIT Press.

Tye, M. (2003) *Consciousness and Persons: Unity and Identity*, Cambridge, MA: MIT Press.

Vosgerau, G., Schlicht, T. & Newen, A. (2008) Orthogonality of phenomenality and content, *American Philosophical Quarterly*, **45** (4), pp. 309–328.

Wells, M.F., Wimmer, R.D., Schmitt, L.I., Feng, G. & Halassa, M.M. (2016) Thalamic reticular impairment underlies attention deficit in Ptchd1(Y/-) mice, *Nature*, **532** (7597), pp. 58–63. doi: 10.1038/nature17427

Wimmer, R.D., Schmitt, L.I., Davidson, T.J., Nakajima, M., Deisseroth, K. & Halassa, M.M. (2015) Thalamic control of sensory selection in divided attention, *Nature*, **526** (7575), pp. 705–709. doi: 10.1038/nature15398

Paper received July 2022; revised October 2022.