Open and Closed loops: A computational approach to Attention and Consciousness

Sabrina Trapp\textsuperscript{c}, Henning Schroll\textsuperscript{a,b}, Fred H Hamker\textsuperscript{b}

\textsuperscript{a} Bernstein Center for Computational Neuroscience, Charite University Hospital, Philippstraße 13, Haus 6, 10115 Berlin, Germany

\textsuperscript{b} Chemnitz University of Technology, Department of Computer Science, Strasse der Nationen 62, 09107 Chemnitz, Germany

\textsuperscript{c} Max Planck Institute for Human Cognitive and Brain Sciences, Department of Neurology, 04103 Leipzig, Germany

Correspondence:
Prof. Dr. Fred Hamker
Chemnitz University of Technology
Department of Computer Science
Strasse der Nationen 62
Chemnitz, 09107, Germany

Phone: +49 (0)371 531-37875
Fax: +49 (0)371 531-25739
Email: fred.hamker@informatik.tu-chemnitz.de

ACKNOWLEDGMENTS.
This research was supported by a grant from the German Research Foundation [DFG HA 2630/6-1] to Fred Hamker.
Abstract

Within the last years, researchers have proposed a double dissociation between attention and consciousness, based both on empirical and conceptual grounds. However, the elusive nature of these constructs complicates progress in their investigation. We present a framework with which we conceptualize attention and consciousness in computational terms. The concepts are embedded in a biologically inspired architecture, spanning the full arc from stimulus to response. Our architecture assumes a general independence of attention and consciousness, but supposes strong interactions. Furthermore, it addresses the developmental aspect, stressing that these functions have to gradually develop through learning.

Key Words: attention, consciousness, interdependence, computational model, closed loops, open loops

Imagine you are watching Shakespeare’s ‘Hamlet’ in a theater. Afterwards, a friend asks you whether you noticed that the skull was from a sheep. You might reply: “No, I didn’t pay attention to the skull”. At the same time, you could state that you haven’t consciously perceived this little detail. Based on phenomenological experience in our daily life, we rarely distinguish between being in an ‘attentive’ and being in a ‘conscious’ state of mind. The intuitive mesh between these constructs has also influenced scientific conceptualization. It has been assumed that attention acts as a mechanistic precursor for conscious experience (Posner, 1994; Mack and Rock, 1998; Dehaene et al., 2001). Multiple empirical studies seem to support the validity that our ability to detect and become consciously aware of visual changes strongly depends on the involvement of attention (Rensink et al., 1997; Simons and Chabris, 1999; O’Regan et al., 2000; Rensink et al., 2000). Additionally, the neural networks and their mutual connections subserving attention and consciousness seem to closely overlap (Rees and Lavie, 2001). Recently, the relationship between attention and consciousness has sparked new interest. Based upon findings that subjects are able to categorize scenes in the near absence of top-down attention (Li et al., 2002; Reddy et al., 2004; Reddy et al., 2006), it has been proposed that the constructs are less interwoven than previously assumed (Koch and Tsuchiya, 2007; van Boxtel et al., 2010). However, many conceptual problems impede an unequivocal decision between attention and consciousness as independent or meshed constructs. Neither ‘attention’ nor ‘consciousness’ are well-defined, unitary constructs. Thus, the construct of attention can be divided into feature-, object-based and spatial attention (i.e. by the kind of selection), bottom-up and top-down attention, selective and divided, or focused and diffuse attention. Koch and Tsuchiya (2007) already limited their proposal to visual top-down attention. However, it is still a matter of
intensive debate whether and to what extent even visual top-down attention is a unitary construct. Thus, it has been assumed that "relatively independent attentional mechanisms operate within different cognitive subsystems depending on the demands of the current stimuli and tasks." (Woodman et al., 2001). 'Attention' then rather constitutes an umbrella term that characterizes control processes in perception and cognition operating at multiple stages of the system (Kastner and Pinsk, 2004). Similarly, there are at least two debated components of consciousness - phenomenal and access consciousness (Block, 1995). While phenomenal consciousness is commonly defined as the content or 'qualia' which are experienced from a first-person perspective, access consciousness is proposed to refer to information that is made available to the brain’s ‘consumer’ systems, e.g. planning or evaluation systems (Dehaene et al., 2001). It was suggested to consider the phenomenal component as what consciousness is, whereas access refers to what consciousness does (Block, 1995). Thus, the latter is a functional notion and is thus in principle amenable to an emulation in synthetic systems. In contrast, the former is usually referred to as the ‘hard problem’ (Chalmers, 1996) and some have speculated that it is qua natura not a tangible subject for scientific investigations (Nagel, 1974). However, although access consciousness is in principle easier to implement, it remains a rather elusive construct. For example, it was further subdivided into ‘awareness-access’ and ‘broadcasting-access’ (Block, 2007). In order to avoid the risk of ending in a merely “semantic debate” without much practical gain, precise conceptualizations are needed. As computational models require to precisely define each of their underlying processes, it might be of use to have a look at attention and consciousness from a computational perspective. We will here argue that – from such a computational perspective – attention and consciousness might be understood as open- and closed loop processes, respectively. We here address primarily the concept of access consciousness and do not address the issue of the first person perspective.

A computational account of attention and consciousness

From a functional point of view, the purpose of attention is traditionally considered as protecting the brain’s limited capacity from informational overload (Broadbent, 1971) or alternatively as a selection mechanism for action or for parameter specification (Allport, 1987; Neumann, 1990). This is thought to be achieved through filtering of information, and filtering logically implies some form of selection. There is evidence from neurophysiology that selection is implemented through a competitive biasing process, in which processing resources are allocated to relevant stimuli (Desimone and Duncan, 1995). Relevance is either defined by the task (task-driven attention), or by the salience of the stimulus (stimulus-
driven attention). The neural consequences of such a selective mechanism are described on various levels, ranging from fMRI studies that report an enhancement of the BOLD response related to the attended stimulus (Kastner et al., 1999; Hopfinger et al., 2000), EEG studies which reveal increased synchrony of firing in the gamma-frequency-band (Fries et al., 2001), to single cell recordings, which provide evidence for modulation of tuning curves (McAdams and Maunsell, 1999; Womelsdorf et al., 2006). As it seems, the computational goal of consciousness, in contrast, is to provide some kind of interpretative ‘snapshot’ of the current state of the system, making it available to planning stages of the brain (Crick and Koch, 1995, 1998). In this sense, the focus is on already present information and its further processing within the system. The conscious, global available content appears to come along with specific structural properties, such as complexity, i.e. highly integrated and differentiated information (Tononi and Edelman, 1998). In the following, we will build upon these functional considerations and argue that it might be useful to think of attention, beyond the mere visual selection, as a mechanism responsible for providing a selective bias within open cortico-subcortical loops while consciousness might be understood as a process of activating memory and stimulus content within closed-loops.

**General Framework**

In our computational framework we postulate the existence of two kinds of cortico-subcortical loops. Closed loops connect a specific ensemble of cortical cells via subcortical structures back to itself. Open loops connect a cortical ensemble to a different one (which may or may not be in the same cortical region). In general, both closed and open loops will involve the thalamus which is heavily connected with cortex and which is of vital importance for both driving and modulating cortical processing (Sherman and Guillery, 2005). A basic kind of reverberant loop therefore consists of mutually excitatory connections between cortex and thalamus. We propose that the learning of the connectivity in such loops involves additional subcortical structures particularly the basal ganglia (BG), but probably hippocampus or cerebellum as well. The BG are a set of nuclei located in the basal forebrain. They are involved in parallel cortico-basal-ganglia-thalamic loops which are classically divided in three main functional domains, even though finer subdivisions can be made (Alexander et al., 1986): the limbic domain, the associative domain, involved in goal-directed learning, action-outcome association and working memory and the sensorimotor domain, involved in selection of action. We suggest that BG are particularly important in learning to activate such loops but might also remain in charge for on-line control. The hippocampus has been hypothesized to be involved in linking
conceptually and perceptually distinct representations, e.g. for episodic memory and memory recall (Squire, 1992; Cohen and Eichenbaum, 1993) and for novelty dependent learning (Lisman and Grace, 2005). Some hints suggest the cerebellum contributes to supervised learning (Doya, 2000).

Closed Loops

Closed loops (Figure 1) allow information to reverberate within them. Thereby, they provide a structural framework that allows information to remain in the system for relatively long periods of time and at that time being processed extensively. Recently, we developed a computational model that learns to maintain information in cortico-BG-thalamic working memory (WM) loops (Schroll et al., submitted). The intimate relationship between consciousness and working memory has been highlighted by several authors (Jackendoff, 1987; Crick and Koch, 1990; Baars and Franklin, 2003). WM maintenance may be thought of as a process of consciously maintaining information that is not sensorily available to the system in specialized closed-loop systems. The developed computational model suggests that WM is not a mechanism rooted in the local recurrent connectivity of prefrontal cortex but rather mediated by the BG, which can learn to establish closed loops to memorize stimulus content if required for solving a particular task. In a more general understanding, closed-loop (reverberant) processing allows any information to stay in the system long enough to be processed extensively. Thus, we hypothesize that the cortico-BG-thalamocortical system is particularly involved in access consciousness and information integration by activating the relevant limbic, associative or motor loops in time.

The thalamocortical system has long before proposed to participate in consciousness (Tononi and Edelman, 1998; Llinas and Pare, 1991; Llinas et al., 1998). Disruptions of functional interactions between cortex and thalamus have been observed during anesthesia (White and Alkire, 2003) and during vegetative state (Laureys et al., 2000) – both being characterized by a lack of consciousness – which has often been interpreted to show that the thalamocortical system provides the necessary activation for conscious states but not for maintaining the content of consciousness. According to our framework we suggest that the thalamocortical system is involved in determining the content of consciousness, particularly through learning via the BG. Consistent with our proposal, a recent fMRI study has shown anesthesia (propofol) to go along with a decrease in functional connectivity between putamen and other brain regions, while relatively preserving thalamo-cortical connectivity (Mhuircheartaigh et al., 2010). The authors conclude that a disruption of subcortical thalamo-
regulatory systems may precipitate a disruption of thalamo-cortical connectivity.

When stimuli or events are processed, different contents will reverberate within different loops. The perception of color and shape of an object for instance, the feelings when touching it and the associations of memories linked to it will likely be analyzed within separate loops. Different from previous accounts which also stressed the relevance of re-entrant processing for conscious experience - particularly if they are widespread (Lamme, 2006) - we deliberately limit the focus of closed-loop processing relevant for consciousness to those cases where the BG are involved. We do not propose that the BG are the site of the content of consciousness, but that they are fundamentally involved in making the content globally available by their control over thalamo-cortical loops. This control could be either indirect through learning, by progressively interlinking loops together before other areas take over their coordination or it could be directly for on-line control, i.e. when the BG stay in charge for switching thalamo-cortical loops on and off.

If consciousness is implemented by closed-loop reverberant activity, then what are the correlates underlying pre-conscious processing? We propose that pre-conscious processing is likely to derive from cessation of stimulation before stimulus-related activity has once passed through relevant loops. In this situation, the neural trace decays before activity begins to reverberate. Clearly within our framework, this cannot produce any conscious experience. The minimal presentation time for a stimulus to be consciously perceived can therefore be defined by the time needed by a piece of information to once travel through the corresponding closed loops. This idea is supported by neurophysiological studies of masking which revealed that particularly the late response of a masked stimulus is missing or is strongly suppressed (Kovács et al., 1995, Rolls et al., 1999, Thompson and Schall, 2000). Masked priming studies (Dehaene et al., 1998; Neumann and Klotz, 1994; Vorberg et al., 2003) revealed that unconscious rendered stimuli are still processed to a significant degree and are not already canceled in the visual system. These observations can be explained by our framework in that closed-loop processing is an active executive process which requires that the relevant stimulus information is available for a sufficient amount of time to establish the correct connectivity.
Open Loops

Open loops (cf. Figure 2) allow selective biasing of information processing in an unidirectional way. A neurophysiologically motivated framework of attention has stressed the importance of top-down biasing signals for attentional selection (Desimone and Duncan, 1995), but little is known about the source of such biasing signals. Prefrontal cortex is known for its involvement in executive functions and by its anatomical connectivity a good candidate for such biasing signals (Miller and Cohen, 2001). However, this does not explain how such biasing signals can be learned to solve the needs for the task at hand. Recently, a computational framework has been provided for how attentional processes can derive from open-loop functioning (Vitay and Hamker, 2010). In this model, prefrontal task-related representations – via cortico-BG-thalamo-cortical loops – bias a competitive interaction between different stimulus representations within ITC in a top-down manner. Depending on which PFC representations are active at a given time, different ITC representations will receive selective excitation. PFC therefore modulates ITC processing by selectively favoring specific representations. The idea of open-loop modulation can easily be generalized to take place between any two cortical ensembles. In computational terms, we thus conceptualize
attention as a process where a specific set of representations activates another set of representations which ultimately bias processing in sensory areas.

Figure 2: Open loops of different complexities. Cell ensembles are shown as circles. Colors indicate which (of two) representations an ensemble is coding. Open loops allow activity to spread from one set of representations to a different one. (A) Open cortico-cortical and cortico-thalamo-cortical loops. (B) Open cortico-basal ganglio-thalamo-cortical loops. BG: basal ganglia; ITC: inferior temporal cortex; PFC: prefrontal cortex; V4: visual area V4

Interactions among Open and Closed Loops

In our framework, consciousness and attention are disjunct in that we implement them by different computational processes. However, this does not exclude that they interact with and depend upon each other. The formation of new closed loops is dependent upon present sensory input and already activated closed loops. Open-loop activation can precede closed-loop activation in the sense that it already boosts a stimulus representation that is supposed to reverberate in a closed loop. However, open-loop biasing can occur independent of closed-loop reverberations which is in accord with empirical findings showing that attentional processes do not necessarily lead to conscious experiences (Naccache et al., 2002; Ansorge and Neumann, 2005; Kentridge et al., 2008; Tapia et al., 2010).

Learning of Consciousness and Attention

From our perspective, attention and consciousness are no fully built-in properties of the brain. Rather, they evolve as an organism interacts with its environment. Both, attention and
access consciousness are executive processes that have to be learned. It is well known that
the neurotransmitter dopamine mediates learning in the striatum, the input nucleus of BG.
According to our computational framework, structural changes in the striatum together with
the simultaneous learning at other sites enable activation of appropriate loops required to
solve a particular task (Schroll et al., submitted). This provides a link between
consciousness and function (Seth, 2009). For access consciousness the underlying function
might be providing the necessary context information for planning and reasoning. Although
our model (Schroll et al., submitted) is still far away from planning and reasoning, it offers a
computational approach to understand consciousness by its function, particularly for
establishing the transition from reactive responses to behavior based on internal
representations. It has been demonstrated that learning in cortico-BG-thalamo-cortical loops
allows an agent to reorganize its actions and active strategies of WM control to maximize
obtained rewards. Thus, it enables an agent to learn to optimize its decisions based on
present and previous information. Similarly, open cortico-BG-thalamic loops learn to allow
reward-related information to extensively bias processing within sensory areas for attentional
selection (Vitay and Hamker, 2010).

There is indeed mounting evidence for the involvement of BG in learning cognitive tasks
(Packard and Knowlton, 2002). In monkey for example, associative learning in the striatum
precedes learning in the prefrontal cortex (Pasupathy and Miller, 2005). Thus, we suggest
learning of cortico-BG-thalamo-cortical loops to be an essential prerequisite for conscious
and attentional processing.

**Discussion**

We outlined a computational framework within which attention and consciousness can be
understood as open- and closed-loop processes, respectively. Our proposal is based upon
previous functional definitions of attention and consciousness. We consider the functional
domain to be a key to the understanding of these constructs as we here see already good
agreement among existing theories and operationalizations. We offer a combined theoretical
framework in which it becomes evident that while being generally independent processes,
attention and consciousness strongly interact. As outlined in the introduction, there is a huge
body of evidence arguing for an interdependence of these constructs. From a more general
point of view, by providing such a unified framework, we account for the often ignored fact
that “A single system (mind) produces all aspects of behavior. […] Even if the mind has parts,
modules, components, […] they all mesh together to produce behavior.” (Newell, 1990).
Finally, our approach stresses the often-overlooked developmental aspect of attention and
consciousness as it offers an interpretation on how attentional biasing and conscious processing emerge to optimally adapt an agent to its environment.

If we simply equated attention with open loop processing and consciousness with closed loop processing, not much would be gained by our framework. However, open and closed loops refer to mechanisms that can be studied whereas attention and consciousness are only labels which require a consensus about its definition. Thus, we now discuss in how far the concepts of open and closed loops could facilitate the understanding of attention and consciousness.

**Open Loops and Theories of Attention**

It is generally assumed that goal-directed attention requires top-down control (Miller and Cohen, 2001) for biasing competitive interactions (Desimone and Duncan, 1995). While several computational models rely on such principles, perhaps one of the most large-scale systems-level models of attention predicts that attention emerges by the interactions of different brain areas, in this model area V4, the inferior temporal cortex (IT), and prefrontal areas including the frontal eye field (FEF) guided by top-down signals allowing for goal directed visual perception (Hamker, 2005a; 2005b, Zirnsak and Hamker, 2006). While this model also shows reentrant processing, e.g. from IT to V4 and from FEF to V4, we do not consider such reentrant loops as closed loops for conscious processing. These loops rather enforce binding and ensure that different brain areas process different aspects of the same content (or physical object in the world). However, top-down signals that have been simply associated to WM in this model, enable control and recruit visual perception for the goals of a task. Vitay and Hamker (2010) have shown that the BG can learn, dependent on a cue, to perform delayed match-to-sample and delayed pair-association tasks. Initially, an object is shown followed by the cue. Then a choice display containing two objects is presented. Dependent on the cue, reward is given if either the same object is chosen or a pair-object. Through dopamine mediated learning, activity in prefrontal cortex allows to recruit a cortico-basal ganglia-thalamocortical loop to provide the correct top-down signal for this task. If we generalize from this computational study, attention is no more than activating the proper top-down signals for visual guidance in an open-loop, let it be a spatial position, a feature, a more high-level feature (or object) or any other pattern that is encoded by visual neurons.
Closed Loops and Theories of Consciousness

In our framework, consciousness is conceptualized as closed-loop processing. Reentrant processing itself is not novel and is considered crucial in most theoretical accounts of consciousness (Edelman, 1989; Grossberg, 1999; Di Lollo et al., 2000; Dehaene et al., 2003; Crick and Koch, 2003; Edelman et al., 2011). For example, it has been suggested to use the feedforward-feedback dichotomy to understand the difference between unconscious and conscious vision (Lamme and Roelfsema, 2000). A central tenet of this proposal is that recurrent processing within the visual cortex is thought to be sufficient for phenomenal consciousness. We fundamentally agree with the basic idea. However, without the specification of the underlying neuroanatomical and functional architecture, recurrent processing remains a necessary, but by no means sufficient correlate of consciousness.

Little is also known about the mechanisms involved in the transition from unconscious to conscious processing. Rather than a threshold mechanism, we propose an active executive process that has to be learned in advance before it can be applied in order to establish reverberating loops. Thus, despite having much in common with previous theories, we emphasize particularly the role of BG in the thalamocortical network and the role of learning.

Conclusion

One major problem of consciousness research so far has been that consciousness and attention are not well defined. By suggesting open and closed loops to differentiate between these constructs, we here provide more precise computational conceptualizations. As such, this framework could also lead to some redefinitions. Without going into detail, so called attentional phenomena such as change blindness, attentional blink and inattentional blindness could be perhaps better described by failures in closed loop processing than in terms of open loops, but future studies particularly devoted to each of these paradigms must address this in more detail.

We might overemphasize the role of BG in thalamocortical circuits. However, previous theories have been very general which makes them difficult to falsify and at the same time provided little guidance in experimental research due to their inherent generality. We have developed a patchwork of computational models (Hamker, 2005a; Hamker 2005b; Vitay and Hamker, 2010; Schroll et al., submitted) which, taken together, provide a framework of attention and consciousness in terms of learning cortico-BG-thalamocortical open and closed loops and attentive selection. This framework could provide guidance to
experimentalists and allows for computational simulations of experiments, even for the
danger that parts of our models will require a revision. We admit that our framework is far
from being complete, but hope, in the style of Warren McCulloch, the reader will not bite our
fingers, but look where we are pointing.

References

circuits linking the basal ganglia and cortex. Annual Review of Neuroscience, 9:357-381.
Allport, A (1987) Selection for action: Some behavioral and neurophysiological considerations of
attention and action. In H. Heuer & A. F. Sanders (Eds.), Perspectives on perception and
action (pp. 395–419). Lawrence Erlbaum Associates.
Ansorge U, Neumann O (2005) Intentions determine the effect of invisible metacontrast-masked
primes: Evidence for top-down contingencies in a peripheral cuing task. J Exp Psychol
Human 31:762-777.
Beh Sci 33:86-86.
Sci 7:166-172.
Press.
MIT Press.
Crick F, Koch C (1990) Towards a neurobiological theory of consciousness. Seminars in the
Neurosciences 2:263-275.
123.
Dehaene S, Sergent C, Changeux JP (2003) A neuronal network model linking subjective reports and
objective physiological data during conscious perception. P Natl Acad Sci USA 100:8520-
8528.
Neuroscience, 18:193-222.
psychophysics of reentrant visual processes. Journal of experimental psychology General
129:481-507.
Doya K (1999) What are the computations of the cerebellum, the basal ganglia and the cerebral
8:1-44.


Rensink RA, O'Regan JK, Clark JJ (1997) To see or not to see: The need for attention to perceive changes in scenes. Psychol Sci 8:368-373.