Make the Unconscious Explicit to Boost the Science of Consciousness

Johan Eriksson [1,2]*, Aurelie Fontan [1,2] and Tiziana Pedale [1,2]

[1] Umeå Center for Functional Brain Imaging (UFBI), Umeå University, Umeå, Sweden, [2] Physiology Section, Department of Integrative Medical Biology, Umeå University, Umeå, Sweden

INTRODUCTION

If we are to understand the how's and why's of conscious processing we need to compare it with unconscious processing and see how they differ (Baars, 1988). An accurate characterization of unconscious processing is therefore crucial for continued progression in the science of consciousness. The scope and limits of unconscious processing have been extensively investigated for more than a century (see e.g., Kouider and Dehaene, 2007, for a historical perspective), but a consensus has yet to emerge. While some argue that unconscious processing is very limited (e.g., Holender and Duscherer, 2004), others have proposed that anything that can be done consciously can also be done unconsciously (Hassin, 2013; but see Hesselmann and Moors, 2015). The controversy continues in part for methodological/technical reasons, for example that it is non-trivial to present a stimulus below the limen of consciousness while maintaining the possibility for the brain to process the information. We argue here that this aspect is particularly relevant to consider when evaluating the possibility of unconscious high-level cognitive functions: if a certain level of processing of the input stimulus cannot be verified to start with, a lack of related high-level cognition would be expected. Such situation provides little evidence against the possibility of a particular type of unconscious high-level cognition. We also argue that evidence against particular unconscious high-level functions should be framed in an experimental context where at least the conscious analog can be demonstrated—otherwise the evidence is again weak. Lastly, we note that an explicit characterization of the unconscious is crucial also in research on the neural correlates of consciousness.

APPROPRIATE BENCHMARKS INCREASE THE WEIGHT OF EVIDENCE AGAINST HIGH-LEVEL UNCONSCIOUS COGNITION

According to several studies, unconscious processing can occur at high perceptual (see Kouider and Dehaene, 2007) and cognitive levels (see Soto and Silvanto, 2014). However, other studies have failed to provide support for high-level unconscious processing (e.g., Schlossmacher et al., 2017; Biderman and Mudrik, 2018). Demonstrating unconscious processing at *any* level is non-trivial, in the sense that it cannot be taken for granted that a stimulus has been registered and processed simply because it has been "unconsciously" presented to the participant; it may be that the signal is too weak to be processed at all. Different techniques to render stimuli unconscious seem to allow processing at different levels (Breitmeyer, 2015). For example, Fahrenfort et al. (2017) found that the attentional blink allowed perceptual integration of unseen Kanizsa inducers, whereas backward masking did not.

There are many possible reasons for null findings (e.g., lack of power, methodological confounds, inappropriate choice of paradigm) but in the context of unconscious high-level cognition, attempts should be made to at least demonstrate unconscious processes at some relevant level to substantiate claims that a particular high-level function cannot be implemented unconsciously. For example,

Schlossmacher et al. (2017) investigated whether facial expressions of emotions could be processed unconsciously during continuous flash suppression. Despite high power, individual adjustment of stimulus contrast, and concurrent measurement of brain activity (which may pick up processes not evident in behavior), they found no difference between neutral and positive/negative facial expressions. Crucially, they did not include a low-level control condition (e.g., scrambled faces, or other appropriate catch trials, see Moors et al., 2019), and it is thus unclear if the null effects were due to an inability to process facial expressions as such, or if the experimental setup did not allow for lower-level processing either, an issue which they themselves discuss extensively (see also Sterzer et al., 2014). They do show some signal change during unconscious trials (Figure 4 in Schlossmacher et al., 2017), but without an appropriate reference condition it is unclear if this change is due to unconscious processing, anticipation, buildup from the Mondrian patterns used during the continuous flash suppression procedure, or something else.

Similarly, Biderman and Mudrik (2018) assessed unconscious object-scene integration using a masking procedure combined with priming in an attempt to replicate previous findings (Mudrik et al., 2011; Mudrik and Koch, 2013). They evaluated whether an unconscious prime including either a congruent or an incongruent object, could influence behavioral performances to categorize a target scene as congruent or incongruent. They did not observe any priming effect and thus claimed no evidence for unconscious integration of objects and scenes. However, as for Schlossmacher et al., Biderman and Mudrik did not include a reference condition such that lower-level processing could be verified. They did include a "visibility post-test" where the same stimulus set was presented either upright or inverted. Here, participants could perform better than chance, which the authors interpreted as indicating partial awareness, in accordance with objective criteria for unconscious perception. Alternatively, one could argue that the upright/inverted task requires a lower level of processing compared to the object-scene integration task and better-than-chance performance could, combined with the subjective measure used in parallel by Biderman and Mudrik, also be interpreted as indicating lower-level unconscious processing. Inclusion of a dedicated reference condition to specifically probe low-level unconscious processing (e.g., a condition where target and prime are identical, to benchmark priming) within the same experiment could have provided valuable information in this regard.

In an earlier attempt to replicate unconscious object-scene integration Moors et al. (2016) also did not find support for integration, but demonstrated a correlation between the time of breaking suppression and low-level image properties. Had there been evidence also of integration, the correlation with low-level image properties could have turned out to be problematic because of the risk of confounding integration with non-relevant image properties (Moors et al., 2019). However, in the context of null support for integration, the correlation with specific image properties could be considered evidence that at least the image content was processed unconsciously. Are the findings by Moors et al. (2016) thereby sufficient to claim that object-scene

integration cannot be performed unconsciously? Ideally, for the specific question of object-scene *integration*, one would like to have seen evidence not only of any form of unconscious image processing, but for object identification and scene identification specifically—only then would it be possible to verify/refute integration of the two. This is a tall order indeed, but this issue illustrates that what constitute an "appropriate benchmark" or "relevant processing level" may differ depending on the research question.

A further point to consider when investigating higher-order processing of unconscious information is if the experimental design can elicit the conscious analog of the process of interest. For example, in a recent magnetoencephalography (MEG) working-memory study, Trübutschek et al. (2019) used a masking procedure to present a barely visible square in different locations, with 20% of the trials being "target absent" trials (which constitutes a reference condition to probe also low-level unconscious processing, as discussed above). The task consisted of maintaining the location of the target square and, after an instruction cue, indicate the original position or mentally rotate the remembered position clockwise or anticlockwise. Thus, on rotation trials the participants were required to manipulate the content held in short-term memory, thereby making it possible to test whether a key feature of working memory (i.e., manipulation) also can be performed unconsciously. The behavioral results revealed better-than-chance performance on unconscious trials, supporting the possibility to maintain and manipulate unconscious information. Neural markers of conscious processing were revealed during the maintenance of conscious compared with unconscious and absent trials, but no difference was evident between unconscious and absent trials, suggesting "activity silent" memory storage (Trübutschek et al., 2017; but see Bergström and Eriksson, 2014, 2018; King et al., 2016). Crucially, neural activity after the presentation of the instruction cue (i.e., during manipulation) revealed no difference between conscious or unconscious vs. absent trials. These results were interpreted as efficient short-term unconscious storage, but a conscious reinstatement of the remembered position to allow the subsequent manipulation of the (until then) unconscious information.

However, the fact that there was no MEG-signal difference between the conscious and absent conditions gives room for an alternative interpretation, namely that the experimental paradigm was suboptimal for characterizing manipulation-related activity. Specifically, it seems that the absent trials required a similar load on relevant cognitive processing as conscious (and unconscious) trials, making the neural markers of manipulation indistinguishable between conditions. Since there was no marker for conscious manipulation of working-memory content, it is not surprising that there was none for unconscious manipulation either. Trübutschek et al. may of course be right in their interpretation, but since there was no explicit demonstration of manipulation-related brain activity at all, the conclusion that unconscious memory representations need to be recast as conscious representations to be manipulable remains speculative.

NEURAL CORRELATES OF CONSCIOUS AND UNCONSCIOUS PROCESSING

An explicit characterization of unconscious processes is also highly desirable in research on the neural correlates of consciousness (NCC). A consensus has been building over the last decade, stating that many traditional paradigms risk conflating the actual NCC with pre-requisite (e.g., attention) and consequent (e.g., reporting) processes (e.g., Aru et al., 2012; Koch et al., 2016). Accordingly, efforts have been made to reduce influence from such confounds, for example through the development of no-report paradigms (see Tsuchiya et al., 2015). However, it remains common practice to assume unconscious processing rather than to explicitly demonstrate it. Unless the neural correlates of *un*conscious processes (NCU) are explicitly characterized and compared with conscious processes, any statement of a NCC will remain hypothetical. Unfortunately, the NCU is many times left uncharted.

For example, when using binocular rivalry (or other bistable stimuli) the experimental rationale is usually that neural activity that follows the shifts in conscious perception rather than the timing of the stimulus itself (which is constant) will mark a NCC (an assumption previously made also by one of the authors). However, not all neural processes that correlate with perceptual switching need to be related to consciousness, as noted previously (Aru et al., 2012; De Graaf et al., 2012; Blake et al., 2014; Giles et al., 2016). Indeed, it was recently demonstrated that binocular rivalry can occur also for unconscious stimuli (Zou et al., 2016). Although there are examples where activity related to the suppressed stimulus is also characterized (e.g., Fries et al., 1997; Tononi et al., 1998), there are other techniques to dissociate stimulus parameters and conscious perception than bistable stimuli that may be better suited to investigate the NCC, for example by using threshold stimuli, manipulations of attention, blindsight, etc. (see e.g., Kim and Blake, 2005; Giles et al., 2016).

A similar argument can be made for the nowadays relatively common approach of manipulating the state of consciousness—usually referring to the level of wakefulness—to investigate the NCC (see Macdonald et al., 2015). The manipulation of the state of consciousness has brought new insight into the NCC by comparing the cognitive and/or brain changes associated with different levels of wakefulness (e.g., sedation vs. no sedation; Boveroux et al., 2010), or during conscious-to-unconscious state transitions (e.g., Cavanna et al., 2018). However, one crucial factor to take into account for the investigation of the NCC is that the manipulation of the global state of consciousness will not only change conscious processes but might also alter several other brain functions. Thus, one cannot rule out that a reduced level of wakefulness could also affect unconscious processing. Until the relationship and/or the differences between conscious and unconscious states are better understood, it seems prudent to control for effects that are not specific to conscious processing (Koch et al., 2016). To be able to assert a neural correlate as related specifically to conscious processing, one could manipulate both content (awareness) and state of consciousness (wakefulness) within the same experiment (Overgaard and Overgaard, 2010; Bachmann and Hudetz, 2014).

CONCLUDING REMARKS

Progress in any field of science hinges on reproducible findings, no less so for a science of consciousness. Arguably, throughout the history of research on unconscious processes there has been more focus on possible false positive findings compared to false negatives. However, true progress requires a handle on both. Here we have argued that extra considerations should be taken in relation to research on unconscious higher-level cognition, such that the weight of evidence from a null finding will increase greatly if presented in a context of appropriate benchmarks. This includes verification that the experimental paradigm can and do support basic processes required for the relevant higher-level process. For example, when studying high-level integration or high-level perceptual processing, a relevant benchmark would be to first show that the information to be integrated/processed is at least discriminated by the brain. Due to the non-trivial nature of implementing an unconscious stimulus presentation, such basic-level processing should not be taken for granted and appropriate verification should be included in each experiment.

By explicitly demonstrating basic-level unconscious processes, the risk of false negative findings can be reduced. We have also argued that there is much to gain by explicitly demonstrating neural correlates of unconscious processes, which can reduce the risk of false positive NCCs. Much has previously been written on the issue of making conscious and unconscious conditions comparable. It remains unclear if this is at all possible, but we here advocate that *at the very least*, neural correlates of unconscious processes should be explicitly demonstrated to, through comparison, substantiate a NCC. Unfortunately, this is not common practice when using bistable stimuli, or when manipulating the state of consciousness.

In conclusion, we believe that progress will increase in the science of consciousness if unconscious processes are made explicit, in the sense that relevant data (behavioral or neural) are explicitly reported to support what are currently all too often implicit background assumptions.

AUTHOR CONTRIBUTIONS

JE, AF, and TP wrote the manuscript.

FUNDING

REFERENCES

Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 737–746. doi: 10.1016/j.neubiorev.2011.12.003

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Bachmann, T., and Hudetz, A. G. (2014). It is time to combine the two main traditions in the research on the neural correlates of consciousness: C=LxD. *Front. Psychol.* 5:940. doi: 10.3389/fpsyg.2014.00940

Bergström, F., and Eriksson, J. (2014). Maintenance of non-consciously presented information engages the prefrontal cortex. *Front. Hum. Neurosci.* 8:938. doi: 10.3389/fnhum.2014.00938

Bergström, F., and Eriksson, J. (2018). Neural evidence for non-conscious working memory. *Cereb. Cortex* 28, 3217–3228. doi: 10.1093/cercor/bhx193

Biderman, N., and Mudrik, L. (2018). Evidence for implicit — but not unconscious — processing of object-scene relations. *Psychol. Sci.* 29, 266–277. doi: 10.1177/0956797617735745

Blake, R., Brascamp, J., and Heeger, D. J. (2014). Can binocular rivalry reveal neural correlates of consciousness? *Philos. Transact. R. Soc. B* 369:20130211. doi: 10.1098/rstb.2013.0211

Boveroux, P., Vanhaudenhuyse, A., Bruno, M.-A., Noirhomme, Q., Lauwick, S., Luxen, A., et al. (2010). Breakdown of within- and between-network resting state functional magnetic resonance imaging connectivity during propofol-induced loss of consciousness. *Anesthesiology* 113, 1038–1053. doi: 10.1097/ALN.0b013e3181f697f5

Breitmeyer, B. G. (2015). Psychophysical "blinding" methods reveal a functional hierarchy of unconscious visual processing. *Consc. Cogn.* 35, 234–250. doi: 10.1016/j.concog.2015.01.012

Cavanna, F., Vilas, M. G., Palmucci, M., and Tagliazucchi, E. (2018). Dynamic functional connectivity and brain metastability during altered states of consciousness. *NeuroImage* 180, 383–395. doi: 10.1016/j.neuroimage.2017.09.065

De Graaf, T. A., Hsieh, P. J., and Sack, A. T. (2012). The "correlates" in neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 191–197. doi: 10.1016/j.neubiorev.2011.05.012

Fahrenfort, J. J., Van Leeuwen, J., Olivers, C. N. L., and Hogendoorn, H. (2017). Perceptual integration without conscious access. *PNAS* 114, 3744–3749. doi: 10.1073/pnas.1617268114

Fries, P., Roelfsema, P. R., Engel, A. K., König, P., and Singer, W. (1997). Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *PNAS* 94, 12699–12704. doi: 10.1073/pnas.94.23.12699

Giles, N., Lau, H., and Odegaard, B. (2016). What type of awareness does binocular rivalry assess? *Trends Cogn. Sci.* 20, 719–720. doi: 10.1016/j.tics.2016.08.010

Hassin, R. R. (2013). Yes it can: on the functional abilities of the human unconscious. *Persp. Psychol. Sci.* 8, 195–207. doi: 10.1177/1745691612460684

Hesselmann, G., and Moors, P. (2015). Definitely maybe: can unconscious processes perform the same functions as conscious processes? *Front. Psychol.* 6:584. doi: 10.3389/fpsyg.2015.00584

Holender, D., and Duscherer, K. (2004). Unconscious perception: the need for a paradigm shift. *Percept. Psychophys.* 66, 872–881. doi: 10.3758/bf03194980

Kim, C.-Y., and Blake, R. (2005). Psychophysical magic: rendering the visible "invisible". *Trends Cogn. Sci.* 9, 381–388. doi: 10.1016/j.tics.2005.06.012

King, J. R., Pescetelli, N., and Dehaene, S. (2016). Brain mechanisms underlying the brief maintenance of seen and unseen sensory information. *Neuron* 92, 1122–1134. doi: 10.1016/j.neuron.2016.10.051

Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/nrn.2016.22

Kouider, S., and Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. *Philos. Trans. R. Soc. London.* 362, 857–875. doi: 10.1098/rstb.2007.2093

Macdonald, A. A., Naci, L., Macdonald, P. A., and Owen, A. M. (2015). Anesthesia and neuroimaging: investigating the neural correlates of unconsciousness. *Trends Cogn. Sci.* 19, 100–107. doi: 10.1016/j.tics.2014.12.005

Moors, P., Boelens, D., Overwalle, J., Van, and Wagemans, J. (2016). Scene integration without awareness : no conclusive evidence for processing scene congruency during continuous flash suppression. *Psychol. Sci.* 27, 945–956. doi: 10.1177/0956797616642525

Moors, P., Gayet, S., Hedger, N., Stein, T., Sterzer, P., van Ee, R., et al. (2019). Three criteria for evaluating high-level processing in continuous flash suppression. *Trends Cogn. Sci.* 23, 267–269. doi: 10.1016/j.tics.2019.01.008

Mudrik, L., Breska, A., Lamy, D., and Deouell, L. Y. (2011). Integration without awareness: expanding the limits of unconscious processing. *Psychol. Sci.* 22, 764–770. doi: 10.1177/0956797611408736

Mudrik, L., and Koch, C. (2013). Differential processing of invisible congruent and incongruent scenes: a case for unconscious integration Liad Mudrik. *J. Vision* 13, 1–14. doi: 10.1167/13.13.24

Overgaard, M., and Overgaard, R. (2010). Neural correlates of contents and levels of consciousness. *Front. Psychol.* 1, 1–3. doi: 10.3389/fpsyg.2010.00164

Schlossmacher, I., Junghöfer, M., Straube, T., and Bruchmann, M. (2017). No differential effects to facial expressions under continuous flash suppression: an event-related potentials study. *NeuroImage* 163, 276–285. doi: 10.1016/j.neuroimage.2017.09.034

Soto, D., and Silvanto, J. (2014). Reappraising the relationship between working memory and conscious awareness. *Trends Cogn. Sci.* 18, 520–525. doi: 10.1016/j.tics.2014.06.005

Sterzer, P., Stein, T., Ludwig, K., Rothkirch, M., and Hesselmann, G. (2014). Neural processing of visual information under interocular suppression: a critical review. *Front. Psychol.* 5:453. doi: 10.3389/fpsyg.2014.00453

Tononi, G., Srinivasan, R., Russell, P. D., and Edelman, G. M. (1998). Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. *PNAS* 95, 3198–3203. doi: 10.1073/pnas.95.6.3198

Trübutschek, D., Marti, S., Ojeda, A., King, J. R., Mi, Y., Tsodyks, M., et al. (2017). A theory of working memory without consciousness or sustained activity. *ELife* 6:e23871. doi: 10.7554/eLife.23871

Trübutschek, D., Marti, S., Ueberschär, H., and Dehaene, S. (2019). Probing the limits of activity-silent non-conscious working memory. *PNAS* 116, 14358–14367. doi: 10.1073/pnas.1820730116

Tsuchiya, N., Wilke, M., Frässle, S., and Lamme, V. A. F. (2015). No-report paradigms: extracting the true neural correlates of consciousness. *Trends Cogn. Sci.* 19, 757–770. doi: 10.1016/j.tics.2015.10.002

Zou, J., He, S., and Zhang, P. (2016). Binocular rivalry from invisible patterns. *Proc. Natl. Acad. Sci. U.S.A.* 113, 8408–8413. doi: 10.1073/pnas.1604816113