# Disorders of artificial awareness

Thomas Parr[1]*, Danijar Hafner[2], and Karl J Friston[1]

[1] Wellcome Centre for Human Neuroimaging, Institute of Neurology, University College London, WC1N 3BG, UK.
[2] Intelligent Systems Group, Department of Computer Science, University College London, WC1E 6BT, UK
thomas.parr.12@ucl.ac.uk, mail@danijar.com, k.friston@ucl.ac.uk

**Abstract.** The study of perceptual awareness in biology often relies upon the study of clinical conditions with either absent or abnormal awareness. In this article, we argue that the same approach may be fruitful in investigating artificial consciousness. To illustrate this, we draw upon recent examples in which disorders of awareness have been induced in artificial systems. Specifically, we call upon the induction of hallucinatory phenomena, and upon visual neglect: a classical disorder of awareness that manifests as a disruption of the action-perception cycle. The key ideas we seek to emphasise from these are the presence of an internal model that generates perceptual content, and the capacity to actively engage with the sensorium.

**Keywords:** Active inference; Awareness; Hallucinations; Visual neglect; Generative models

## 1 Introduction

Disorders of awareness have formed the basis for neuropsychological investigations into aspects of conscious experience in humans [1]. Part of the reason this approach has been so popular is that, while awareness is difficult to define, it is often easy to recognise its absence. In this article, we suggest that an analogous approach could yield new insights in artificial consciousness research. To illustrate this, we describe two recent accounts of abnormal perception induced in (simulated) artificial systems. Crucially, these replicate phenomena observed in human disorders in which awareness is either impaired or augmented. First, we outline a computational account of visual hallucinations that depends upon a failure of sensory systems to correct internally generated perceptual content. Second, we discuss the importance of action in sampling the world, and the consequences of its failure in a synthetic version of visual neglect.

To formalise the concepts above, it is useful to frame them in terms of active inference [2]. This is a way to describe perception and action that appeals to the minimisation of variational free energy, which depends upon a generative model that describes how a (living or artificial) system believes their sensory data are generated, and upon beliefs about the current state of the world [3, 4]. There are two ways in which free

energy may be minimised. The first is by optimising posterior beliefs such that they become more consistent with sensory data (i.e. inference). The second is by acting to change sensory data such that they conform to current beliefs. Combining the two, it becomes possible to infer future plans of action that will yield sensory data that minimise expected free energy [5]. Intuitively, we can think of this as a scientific endeavour in which we use current sensory data to test hypotheses about their causes (minimising free energy), and use the resulting inferences to plan future experiments to gather more data (minimising expected free energy).

## 2 Disorders of Awareness

Hallucinations offer an interesting perspective on awareness, as they represent awareness of fictitious perceptual content. Given the absence of supportive sensory data, this implies awareness can depend purely upon internally generated percepts [6]. To simulate this phenomenon, we constructed a generative model that included prior beliefs about abstract visual objects, and about the scene in which they were embedded [7]. By equipping the model with beliefs about the reliability of sensory data, we found that confidence in sensory data was down-weighted when data was inconsistent with prior beliefs. In some instances, this was sufficient to release perceptual inference from the constraints of these data, leading to hallucinatory phenomena (i.e. false positive inferences) that preserved the internal consistency of the scene, consistent with many biological hallucinations [8]. This emphasises the importance of data in modulating perceptual awareness, and the need to seek out informative, high quality, sensations.

Pursuing the simile of the brain, or an artificial equivalent, as a scientist, we turn to planning as a process of experimental design, and what this means for evaluating the 'goodness' of a plan. The best experiments are those that bring about a large change in beliefs. This has been formalised in the notion of information gain (or expected free energy), which has a long history in experimental design [9], and has been employed to understand salience in visual search [10]. The imperative to perform uncertainty-resolving perceptual experiments may also be leveraged to account for a classic disorder of awareness; visual neglect. Neglect is a classic neuropsychological disorder of visual awareness in which one side of space is ignored [11] that can manifest as a poverty of saccadic sampling (perceptual experimentation) in this hemifield [12]. If we are very confident in beliefs about variables associated with specific regions of space, experiments (e.g. eye movements) that obtain data at these locations afford very little information gain. By setting prior beliefs about mappings from causes (fixations) to consequences (visual data) to be highly confident on one side of space, we were able to reproduce the behavioural phenomenology of visual neglect [13].

## 3 Conclusion

In brief, we have summarised two instances in which disorders of perception have been replicated in synthetic systems. Each relies upon a failure to incorporate sensory data into perceptual inference; either due to a failure to use these data to constrain internally generated content, or a failure to acquire it in the first place. These synthetic disorders of awareness are highly consistent with philosophical perspectives on consciousness as a process of inference [14, 15]. Artificial consciousness, like its biological homologue, may benefit from the study of its disorders. While there may be many different approaches to develop conscious systems [16], a bidirectional engagement with their environment must be a key feature. Consequently, understanding the generation of spurious perceptual content, and the absence of awareness characteristic of neglect disorders, offers a new perspective on the requirements for synthetic conscious awareness in terms of the generative models that underwrite inferential procedures.

## Acknowledgements

TP is supported by the Rosetrees Trust (Award Number 173346). KJF is a Wellcome Principal Research Fellow (Ref: 088130/Z/09/Z).

## References

1. Rees, G., G. Kreiman, and C. Koch, *Neural correlates of consciousness in humans.* Nature Reviews Neuroscience, 2002. **3**: p. 261.
2. Friston, K.J., et al., *Action and behavior: a free-energy formulation.* Biological Cybernetics, 2010. **102**(3): p. 227-260.
3. Dayan, P., et al., *The Helmholtz machine.* Neural computation, 1995. **7**(5): p. 889-904.
4. Beal, M.J., *Variational algorithms for approximate Bayesian inference*. 2003, University of London United Kingdom.
5. Friston, K., et al., *Active inference and epistemic value.* Cognitive Neuroscience, 2015. **6**(4): p. 187-214.
6. Adams, R.A., et al., *The Computational Anatomy of Psychosis.* Frontiers in Psychiatry, 2013. **4**: p. 47.
7. Parr, T., et al., *Precision and false perceptual inference.* Front. Integr. Neurosci., 2018.
8. Collerton, D., E. Perry, and I. McKeith, *Why people see things that are not there: A novel Perception and Attention Deficit model for recurrent complex visual hallucinations.* Behavioral and Brain Sciences, 2005. **28**(6): p. 737-757.
9. Lindley, D.V., *On a Measure of the Information Provided by an Experiment.* Ann. Math. Statist., 1956. **27**(4): p. 986-1005.

10. Itti, L. and P. Baldi, *Bayesian surprise attracts human attention.* Advances in neural information processing systems, 2006. **18**: p. 547.
11. Halligan, P.W. and J.C. Marshall, *Neglect of Awareness.* Consciousness and Cognition, 1998. **7**(3): p. 356-380.
12. Fruhmann Berger, M., L. Johannsen, and H.-O. Karnath, *Time course of eye and head deviation in spatial neglect.* Neuropsychology, 2008. **22**(6): p. 697-702.
13. Parr, T. and K.J. Friston, *The Computational Anatomy of Visual Neglect.* Cerebral Cortex, 2017: p. 1-14.
14. Hohwy, J., *Attention and Conscious Perception in the Hypothesis Testing Brain.* Frontiers in Psychology, 2012. **3**: p. 96.
15. Gregory, R.L., *Perceptions as Hypotheses.* Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 1980. **290**(1038): p. 181.
16. Gamez, D., *Human and Machine Consciousness*. 2018: Open Book Publishers.