SUPPORTING INFORMATION

Feeling It: A Higher-Order Theory of Emotional Consciousness.
Joseph LeDoux and Richard Brown

Table of Contents **Page Number**

Box 1: Defining Cortical vs. Subcortical Circuits 2

Box 2: The Consciousness Patients: How Neurological Cases Helped Revive Interest in Consciousness as a Scientific Topic 2

Box 3: Measuring Introspective Consciousness Using Verbal and other Reporting Methods in Humans and Other Animals 3

Box 4: Nonconscious Cognition (Especially Nonconscious Working Memory) 7

Box 5: Rich versus Sparse views on Phenomenology of Perception 10

Box 6: Role of Awareness in First-Order Phenomenal Consciousness 12

Box 7: Higher-Order Theories and their Relation to Other Cognitive Theories of Consciousness 15

Box 8: Phenomenal Consciousness in Higher-order Thought Theories of Consciousness 19

Box 9: Relational versus Non-Relational Higher-Order Theories 19

Box 10: Research Questions 21

BOX 1: DEFINING CORTICAL VS. SUBCORTICAL CIRCUITS.

For the purposes of this article we distinguish "cortical" and "subcortical" circuits of the cerebral hemispheres in the following way. "Cortical" is used in reference to laminated structures in the outer shell (i.e. the cortex) of cerebral hemispheres. Particular emphasis is placed on cortical regions that contribute to sensory processing (areas of visual cortex) and higher cognitive processing (especially lateral and medial prefrontal cortex, but also parietal and insular cortex). "Subcortical" is used to refer to structures that lie within, but underneath the cortical mantle, of the cerebral hemispheres, and that lack prominent or consistent lamination. Circuits centered on the amygdala that detect innate and learned threats and control the expression of defensive behaviors and supporting physiological responses in response to theses stimuli are emphasized.

BOX 2: THE CONSCIOUSNESS PATIENTS: HOW NEUROLOGICAL CASES HELPED REVIVE INTEREST IN CONSCIOUSNESS AS A SCIENTIFIC TOPIC

Consciousness was the number one psychological topic of interest to philosophers and psychologists until the early 20th century[1]. In an effort to make psychology an objective science, observable behavior, rather than inner states of mind knowable only by introspection, came to be emphasized by so-called behaviorists[2, 3]. They felt no need to go inside the "black box." In their view, behavior could be explained in terms of relations between observable stimuli and responses, and neither mental nor neural explanations were needed. Towards mid-century, cognitive science brought the mind back to psychology but not the conscious mind that the behaviorists eliminated; the cognitive mind was more an information processing system, rather that a system of conscious experiences. But by the 1960s, consciousness had begun to make a comeback, inspired in large part by findings from patients who had undergone brain surgery or who had suffered brain injuries.

Observations of split-brain patients, in whom the cerebral commissures were sectioned in order to control epilepsy[4] played a particularly pivotal role in reawakening interest in consciousness as a scientific topic. These studies showed that conscious experiences are inexorably tied to localized processes in the brain—the left hemisphere could verbally describe stimuli that it saw but denied seeing stimuli presented to the right hemisphere[5]. That the right side had processed the stimulus that could not be reported on was indicated by subject's ability to respond non-verbally to the stimulus. This kind of fracture between verbal and non-verbal reporting remains a cornerstone of consciousness research today (see Box 3). In later studies of these patients behavioral responses were triggered from the right hemisphere, and the patient was then asked why he did that. Verbal reports from the left hemisphere explained the behavior in ways that made some sense given what was observed (if the right hemisphere produced a scratching action by the left hand, the left hemisphere said, "I had an itch"). But these were fabrications. Such observations suggested that a role of consciousness is to explain responses generated by non-conscious brain systems[5, 6].

Studies of patients with amnesia resulting from damage to the medial temporal lobe (MTL) also contributed to the revival of interest in consciousness.

MTL amnesia was initially thought to reflect a global loss of memory[7] but later studies suggested that memory deficits were more restricted[8]. Further work made it clear that MTL memory syndrome mainly involved the inability to form new consciously accessible memories. This led to the idea that explicit or conscious memories are anatomically distinct from memories created by non-conscious systems that are involved in sensory processing, motor control, or conditioning[9-12].

Patients with damage to the right parietal cortex exhibit a classic neurological syndrome called unilateral neglect in which they fail to notice stimuli presented in the left visual field[13]. As in split-brain patients, if two stimuli are presented, one in each visual field, only the stimulus in the right visual field is reported. Later studies showed that parietal neglect patients could state whether two stimuli were the same or different, in spite of only being able to report on the identity of the stimulus in the right visual field[14]. Neglect is commonly thought of as an attention deficit[15, 16], a view that contributed to current interest in parietal cortex as part of the cognitive circuits that underlie conscious experiences[17, 18].

A related set of findings came from studies of patients with damage to the right visual cortex. These so-called blindsight patients fail to report on left visual field stimuli but can respond behaviorally to the same stimuli[19-22]. Blindsight findings have often been called upon in studies of the neural basis of consciousness to demonstrate that it is possible for one to respond to visual stimuli without consciously knowing what was seen.

Box 2 References

1. Boring EG (1950) A history of experimental psychology. New York: Appleton-Century-Crofts.
2. Watson JB (1925) Behaviorism. New York: W.W. Norton.
3. Skinner BF (1938) The behavior of organisms: An experimental analysis. New York: Appleton-Century-Crofts.
4. Gazzaniga MS, Bogen JE, Sperry RW (1962) Some functional effects of sectioning the cerebral commissures in man. Proceedings of the National Academy of Sciences USA 48:1765-1769.
5. Gazzaniga MS (1972) One brain - Two minds. American Scientist 60:311-317.
6. Gazzaniga MS, LeDoux JE (1978) The Integrated Mind. New York: Plenum.
7. Scoville WB, Milner B (1957) Loss of recent memory after bilateral hippocampal lesions. Journal of Neurology and Psychiatry 20:11-21.
8. Corkin S (1968) Acquisition of motor skill after bilateral medial temporal lobe excision. Neuropsychologia 6:255-265.
9. Milner B (1965) Memory disturbances after bilateral hippocampal lesions in man. In: Cognitive Processes and Brain (Milner, P. M. and Glickman, S. E., eds) Princeton: Van Nostrand.
10. Squire L (1987) Memory and Brain. New York: Oxford.
11. Schacter DL (1998) Memory and awareness. Science 280:59-60.
12. Eichenbaum H (2002) The cognitive neuroscience of memory. New York: Oxford University Press.
13. Critchley M (1953) The Parietal Lobes. London: Edward Arnold.
14. Volpe BT, LeDoux JE, Gazzaniga MS (1979) Information processing of visual stimuli in an 'extinguised' field. Nature 282:722-724.
15. Heilman KM, Valenstein E, Watson RT (2000) Neglect and related disorders. Semin Neurol 20:463-470.
16. Mesulam MM (1981) A cortical network for directed attention and unilateral neglect. Ann Neurol 10:309-325.
17. Vuilleumier P, Armony JL, Clarke K, Husain M, Driver J, Dolan RJ (2002) Neural response to emotional faces with and without awareness: event-related fMRI in a parietal patient with visual extinction and spatial neglect. Neuropsychologia 40:2156-2166.

18. Rees G, Wojciulik E, Clarke K, Husain M, Frith C, Driver J (2002) Neural correlates of conscious and unconscious vision in parietal extinction. Neurocase 8:387-393.
19. Weiskrantz L, Warrington EK, Sanders MD, Marshall J (1974) Visual capacity in the hemianopic field following a restricted occipital ablation. Brain 97:709-728.
20. Weiskrantz L (1986) Blindsight: A case study and implications. Oxford: Clarendon Press.
21. Cowey A, Stoerig P (1991) The neurobiology of blindsight. Trends Neurosci 14:140-145.
22. Weiskrantz L (1997) Consciousness lost and found: A neuropsychological exploration. New York: Oxford University Press.

BOX 3: MEASURING INTROSPECTIVE CONSCIOUSNESS USING VERBAL AND OTHER REPORTING METHODS IN HUMANS AND OTHER ANIMALS

Assessment of consciousness requires some form of self-reporting[1, 2]. Humans can typically give either a verbal or a nonverbal report of information to which we have introspective access, but cannot provide a verbal report of information that is only processed nonconsciously[3-5]. Verbal self-reports are thus generally agreed to be useful for assessing consciousness, but ineffective in assessing processes that are implicit or unconscious[3-5]. Non-verbal reporting is the only option for assessing non-conscious processing, but is less satisfactory as a measure of consciousness because it can reflect both processes one is aware of and processes that are not introspectively accessible[6]. Fractures between conscious and nonconscious processes by verbal and nonverbal responses have thus played a key role in demonstrating introspective awareness[7-14].

These points are illustrated by findings in studies of patients with blindsight[1, 15-17] or in people with healthy brains tested using subliminal stimulation techniques such as brief exposures[18], visual making[19-26], or continuous flash suppression[27, 28] to impair conscious perception. In these various kinds of studies, people can produce either verbal or nonverbal responses to stimuli presented in free vision but can only respond nonverbally to "unseen" stimuli.

While verbal self-reporting is thus an important tool in consciousness research, it has limitations. It is most suitable for assessing the content of immediate experiences, and its usefulness decreases as the time between the experience and the report increases because distortions and fabrications can enter[3, 4]. It is thus less useful for assessing the motivations underlying some past action, both because of the delay in the report, and also because motivations are often not consciously available and verbalizable[29-34]. Self-report also has limits, when used alone, for assessing self-knowledge about inner experiences, as opposed to the content of experience. Another procedure, post-decision wagering, has been proposed to be an objective measure of consciousness since self-report is not requested[35]. However, since wagering does require introspection, its status as more objective is questionable[5]. Also, post-decision wagering is based on memory and thus potentially suffers from distortions and fabrications. Ratings of one's confidence in perceptual decisions is thought to be helpful as a supplement to self-report in meta-cognitive decision making tasks[36], especially if made concurrently with the experience[5]. Currently, the value of confidence measures vs. verbal reporting as measures of inner awareness is debated[2, 5, 37-39].

Some propose that nonhuman animals have inner experiences, but, lacking language, are simply less facile in their ability to report. But the problem is not so

simply dismissed. Since animals can only give nonverbal reports[1], there is no second reporting method that can distinguish nonconscious from conscious processing. Nonverbal behavior is satisfactory for demonstrating that an animal is conscious in the sense of being awake and responsive to stimuli. It is also sufficient for demonstrating cognitive capacities such as working memory, attention, metacognition, and problem solving ability and other indicators of intelligent behavior[40]. While consciousness depends on cognitive processing, not all cognitive processing, including working memory and metacognition, leads to conscious experience (see main text).

Deciding whether a non-verbal behavior reflects conscious vs. nonconscious cognitive processes requires not only that that the behavior be explainable in terms of conscious processes, but also that nonconscious explanations are inadequate[41]. While there is a new wave of enthusiasm for the idea of animal consciousness[42-47], most of the claims are based on findings showing that the behavior is plausibly explained in terms of consciousness; the alternative hypothesis is less often considered. Since many aspects of human behavior can be accounted for by nonconscious processes[30, 34, 48,49], the most direct way to draw the line between the presence and absence of introspective consciousness is by contrasting verbal and nonverbal reports, which cannot be done in animals. Behavioral evidence for cognitive capacities is not necessarily evidence for conscious awareness in animals since much of cognition occurs nonconsciously.

Two points of clarification need to be made about animal consciousness. First, we do not claim that nonhuman organisms lack inner experiences, but instead claim that because of the difficulty of fracturing conscious and nonconscious processing in nonverbal organisms, the science of consciousness, at least for now, is most fruitfully pursed in humans. Second, given that important aspects of cognitive processing occur nonconsciously in humans, even if animals turn out to be limited to nonconscious processes, this would not mean that they are reflexive "beast machines"[41, 50, 51]. Some animals clearly use sophisticated cognitive processes, many of which fall into the category of the cognitive unconscious[48] to guide complex behaviors.

Box 3 References

1. Weiskrantz L (1997) Consciousness lost and found: A neuropsychological exploration. New York: Oxford University Press.
2. Overgaard M, Sandberg K (2014) Kinds of access: Different methods for report reveal different kinds of metacognitive access. In: The Cognitive Neuroscience of Metacognition (Fleming, S. M. and Frith, C. D., eds), pp 67-86: Springer-Verlag Berlin Heidelberg.
3. Ericcson KA, Simon H (1993) Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press.
4. Wilson TD (1994) THE PROPER PROTOCOL: Validity and Completeness of Verbal Reports. Psychological Science 5:249-252.
5. Seth AK, Dienes Z, Cleeremans A, Overgaard M, Pessoa L (2008) Measuring consciousness: relating behavioural and neurophysiological approaches. Trends Cogn Sci 12:314-321. PMC2767381.
6. LeDoux JE (2015) Anxious: Using the brain to understand and treat fear and anxiety. New York: Viking.
7. Frith C, Perry R, Lumer E (1999) The neural correlates of conscious experience: an experimental framework. Trends in Cognitive Sciences 3:105-114.

8. Jack AI, Shallice T (2001) Introspective physicalism as an approach to the science of consciousness. Cognition 79:161-196.
9. Dennett DC (1991) Consciousness explained. Boston: Little, Brown and Company.
10. Jackendoff R (2007) Language, Consciousness, Culture: Essays on Mental Structure. Cambridge: MIT Press.
11. Naccache L, Dehaene S (2007) Reportability and illusions of phenomenality in the light of the global neuronal workspace model. Behav Brain Sci 30:518-520.
12. Sergent C, Rees G (2007) Conscious access overflows overt report. Behav Brain Sci 30:523-524.
13. Koch C, Tsuchiya N (2007) Phenomenology without conscious access is a form of consciousness without top-down attention. Behav Brain Sci 30:509-510.
14. Nahmias EA (2002) Verbal reports on the contents of consciousness: Reconsidering intropectionist methodology. Psyche 8.
15. Weiskrantz L (1986) Blindsight: A case study and implications. Oxford: Clarendon Press.
16. Cowey A, Stoerig P (1991) The neurobiology of blindsight. Trends Neurosci 14:140-145.
17. Weiskrantz L, Warrington EK, Sanders MD, Marshall J (1974) Visual capacity in the hemianopic field following a restricted occipital ablation. Brain 97:709-728.
18. Lazarus R, McCleary R (1951) Autonomic discrimination without awareness: A study of subception. Psychological Review 58:113-122.
19. Merikle PM (1982) Unconscious perception revisited. Percept Psychophys 31:298-301.
20. Cheesman J, Merikle PM(1986). Distinguishing conscious from unconscious perceptual processes. Can J Psychol 40:343-67.
21. Marcel AJ (1983) Conscious and unconscious perception: experiments on visual masking and word recognition. Cognitive psychology 15:197-237.
22. Greenwald AG, Klinger MR (1990) Visual masking and unconscious processing: differences between backward and simultaneous masking? Mem Cognit 18:430-435.
23. Greenwald AG, Klinger MR, Schuh ES (1995) Activation by marginally perceptible ("subliminal") stimuli: dissociation of unconscious from conscious cognition. J Exp Psychol Gen 124:22-42.
24. Breitmeyer BG, Ogmen H (2006) Visual Masking: Time Slices Through Conscious and Unconscious Vision. Oxford: Oxford University Press.
25. Kouider S, Dehaene S (2007) Levels of processing during non-conscious perception: a critical review of visual masking. Philosophical transactions of the Royal Society of London Series B, Biological sciences 362:857-875. PMC2430002.
26. Macknik SL (2006) Visual masking approaches to visual awareness. Progress in brain research 155:177-215.
27. Yang E, Brascamp J, Kang MS, Blake R (2014) On the use of continuous flash suppression for the study of visual processing outside of awareness. Frontiers in psychology 5:724. PMC4093749.
28. Li HH, Carrasco M, Heeger DJ (2015) Deconstructing Interocular Suppression: Attention and Divisive Normalization. PLoS Comput Biol 11:e1004510. PMC4627721.
29. Nisbett RE, Wilson TD (1977) Telling more than we can know: Verbal reports on mental processes. Psychological Review 84:231-259.
30. Bargh JA, Morsella E (2008) The Unconscious Mind. Perspectives on psychological science: a journal of the Association for Psychological Science 3:73-79. PMC2440575.
31. Banaji MR, Greenwald AG (1995) Implicit gender stereotyping in judgments of fame. J Pers Soc Psychol 68:181-198.
32. Devos T, Banaji MR (2003) Implicit self and identity. Annals of the New York Academy of Sciences 1001:177-211.
33. Phelps EA, O'Connor KJ, Cunningham WA, Funayama ES, Gatenby JC, Gore JC, Banaji MR (2000) Performance on indirect measures of race evaluation predicts amygdala activation. J Cogn Neurosci 12:729-738.
34. Hassin RR, Uleman JS, Bargh JA (eds.) (2005) The new unconscious. New York: Oxford University Press.
35. Persaud N, McLeod P, Cowey A (2007) Post-decision wagering objectively measures awareness. Nature neuroscience 10:257-261.

36. Lau HC, Passingham RE (2006) Relative blindsight in normal observers and the neural correlate of visual consciousness. Proceedings of the National Academy of Sciences of the United States of America 103:18763-18768. PMC1693736.
37. Rosenthal D (2012) Higher-order awareness, misrepresentation and function. Philosophical transactions of the Royal Society of London Series B, Biological sciences 367:1424-1438.
38. Maniscalco B, Lau H (2016) The signal processing architecture underlying subjective reports of sensory awareness. Neurosci Conscious 2016. PMC4972343.
39. Cortese A, Amano K, Koizumi A, Kawato M, Lau H (2016) Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. Nat Commun 7:13669.
40. Smith JD (2009) The study of animal metacognition. Trends Cogn Sci 13:389-396.
41. Heyes C (2008) Beast machines? Questions of animal consciousness. In: Frontiers of Consciousness: Chichelle lectures (Weiskrantz, L. and Davies, M., eds), pp 259-274 Oxford: Oxford University Press.
42. Gross M (2013) Elements of consciousness in animals. Current biology 23:R981-983.
43. Bekoff M (2007) The Emotional Lives of Animals: A leading scientist explores animal joy, sorrow, and empathy - and why they matter. Novato, CA: New World Library.
44. Panksepp J (2012) The Archaeology of Mind: Neuroevolutionary Origins of Human Emotion New York: W. W. Norton & Company.
45. de Waal F (2009) The Age of Empathy: Nature's lessons for a kinder society. New York: Three Rivers Press.
46. Edelman DB, Seth AK (2009) Animal consciousness: a synthetic approach. Trends Neurosci 32:476-484.
47. Low P (2012) Cambridge Declaration on Consciousness in Non-Human Animals. (also by J Panksepp, D Reiss, D Edelman, and B van Swinderen, and C Koch). Originally retrieved on Sept 26, 2013 from http://fcmconferenceorg/Churchill College, University of Cambridge. This link was subsequently removed. A search on Dec 24, 2014, revealed that the document was again available through this link: http://fcmconference.org/img/CambridgeDeclarationOnConsciousness.pdf In: http://fcmconferenceorg/ Churchill College, University of Cambridge.
48. Kihlstrom JF (1987) The Cognitive Unconscious. Science 237:1445-1452.
49. Wilson TD (2002) Strangers to ourselves: Self-insight and the adaptive unconscious. Cambridge, MA: Harvard University Press.
50. Descartes R (1644) Principia philosophiae. Ghent University: apud Ludovicum Elzevirium.
51. Dickinson A (2008) Why a rat is not a beast machine. In: Frontiers of Consciousness (Weiskrantz, L. and Davies, M., eds), pp 275-288 Oxford: Oxford University Press.

BOX 4: NONCONSCIOUS COGNITION (ESPECIALLY NONCONSCIOUS WORKING MEMORY)

Cognitive theories generally assume that information processing occurs non-consciously until it is rendered conscious through processes that involve working memory and attention. The non-conscious processing is often assumed to be sensory in nature. However, considerable evidence has begun to emerge suggesting that cognitive processes that underlie working memory representations involving prefrontal and parietal areas can occur without generating conscious content[1-15].

Block[16] accepts the existence of nonconscious working memory but questions it's capacity. Based on the work of Soto et al[9], he has argued that non-conscious representations in working memory are too weak to account for the performance of subjects in studies of perceptual temporary memory using the partial report paradigm[17, 18]. But in contrast to partial report studies, in which subjects are presented an array of letters or shapes for various intervals and then presented with a cue after the stimulus is no longer available, Soto et al used

masking to degrade the stimulus and prevent awareness and reporting. Masking and other brief stimulus presentations weaken the percept[19, 20]. In a head to head matchup this gives conscious processing a significant edge purely on methodological grounds[21].

Block counters such challenges by citing work[22, 23] suggesting that even when the representations in working memory are robust they do not last long enough to account for existing findings[24]. However, more recent research shows that non-conscious temporary retention involving prefrontal cortex processing can last considerably longer than previously reported, is resistant to distraction, and is related to prospective action[1, 25-27]. These findings thus indicate that nonconscious working memory is long-lasting, robust, and capable of guiding behavior.

But another factor may be at work besides weak percepts. Conscious experience of a stimulus may itself make the underlying lower-order representations more robust[28], what might be loosely called conscious amplification, which might simply be attentional amplification, of lower-order representations.

For these reasons (weak percepts and conscious amplification), findings that working memory capacity is weak in conditions with weak representations should not be used to cast doubt on how robust working memory is when it deals with strong representations (like the kind produced when one consciously experiences the stimulus). To evaluate the capacity/robustness of nonconscious working memory, studies of blindsight patients might be more useful than masking-type procedures since nonconscious stimulus exposures lasting seconds can be used.

These issues can be considered from another perspective. Long-term memory stored via the medial temporal lobe memory system is unconscious until retrieved into working memory.[21] This clearly shows how non-conscious cognitive (as opposed to sensory) information can exist as lower-order non-conscious representation that is re-represented in working memory circuits to render it conscious.

Conscious awareness is thus based on non-conscious processes, and there is no intrinsic reason why all activity in prefrontal cortex must be conscious. In other words, non-conscious representations in prefrontal cortex can be rendered conscious by being re-represented in working memory. Whether the non-conscious representation in prefrontal cortex is "in" working memory or not is a different question. The point is that a theory of conscious awareness is not just about how sensory stimuli come to be experienced. It is about how conscious experiences in a more general sense comes about. Sensory consciousness is an excellent starting point[29] but should not be a limiting factor in a broader theoretical understanding of consciousness.

We thus propose that evidence that processes like attention and working memory are implicated in some task does not show that the task depends on consciousness. Because attention and working memory can operate nonconsciously, these processes, while necessary, are not sufficient for phenomenal experience. Only when that non-conscious information is re-represented in the appropriate way does phenomenal awareness occur[21, 28]. A similar view has recently been expressed by others who note that the phenomenal experience of

information in working memory involves, through additional top-down modulation, the creation of an additional distinct representation of memory content[30, 31]. This leaves open the question of what is the appropriate higher-order representation in working memory. In other words, prefrontal circuits that contribute to consciousness do not constitute a singular unified system[19, 20]. Future research will need to further separate the different cognitive processes contributed to by these circuits, assess their contributions to awareness, and determine whether some contribute to nonconscious representations and others to higher-order representations of these that render the processing conscious.

Box 4 References

1. Soto D, Silvanto J (2014) Reappraising the relationship between working memory and conscious awareness. Trends Cogn Sci 18:520-525.
2. Custers R, Aarts H (2010) The unconscious will: how the pursuit of goals operates outside of conscious awareness. Science 329:47-50.
3. Lau HC, Passingham RE (2007) Unconscious activation of the cognitive control system in the human prefrontal cortex. J Neurosci 27:5805-5811.
4. Del Cul A, Dehaene S, Reyes P, Bravo E, Slachevsky A (2009) Causal role of prefrontal cortex in the threshold for access to consciousness. Brain 132:2531-2540.
5. van Boxtel JJ, Tsuchiya N, Koch C (2010) Consciousness and attention: on sufficiency and necessity. Frontiers in psychology 1:217. PMC3153822.
6. van Gaal S, Lamme VA (2012) Unconscious high-level information processing: implication for neurobiological theories of consciousness. The Neuroscientist 18:287-301.
7. Thakral PP (2011) The neural substrates associated with inattentional blindness. Conscious Cogn 20:1768-1775.
8. Hassin RR, Bargh JA, Engell AD, McCulloch KC (2009) Implicit working memory. Conscious Cogn 18:665-678. PMC2760263.
9. Soto D, Mantyla T, Silvanto J (2011) Working memory without consciousness. Current biology : CB 21:R912-913.
10. Tsuchiya N, Koch C (2009) The relationship between consciousness and attention. In: The Neurology of Consciousness (Laureys, S. and Tononi, G., eds), pp 63-77 New York: Elsevier.
11. Ansorge U, Horstmann G, Scharlau I (2011) Top-down contingent feature-specific orienting with and without awareness of the visual input. Advances in cognitive psychology / University of Finance and Management in Warsaw 7:108-119. PMC3260021.
12. Kiefer M (2012) Executive control over unconscious cognition: attentional sensitization of unconscious information processing. Front Hum Neurosci 6:61. PMC3311241.
13. Bor D, Seth AK (2012) Consciousness and the prefrontal parietal network: insights from attention, working memory, and chunking. Frontiers in psychology 3:63. PMC3298966.
14. Cohen MA, Cavanagh P, Chun MM, Nakayama K (2012) The attentional requirements of consciousness. Trends Cogn Sci 16:411-417.
15. Persaud N, Davidson M, Maniscalco B, Mobbs D, Passingham RE, Cowey A, Lau H (2011) Awareness-related activity in prefrontal and parietal cortices in blindsight reflects more than superior visual performance. NeuroImage 58:605-611.
16. Block N (2011) Perceptual consciousness overflows cognitive access. Trends Cogn Sci 15:567-575.
17. Sperling G (1960) The information available in brief visual presentations. Psychological Monographs 74:1-29.
18. Sligte IG, Scholte HS, Lamme VA (2008) Are there multiple visual short-term memory stores? PLoS One 3:e1699. PMC2246033.
19. Lau H, Rosenthal D (2011) Empirical support for higher-order theories of conscious awareness. Trends Cogn Sci 15:365-373.
20. Lau H, Rosenthal D (2011) The higher-order view does not require consciously self-directed introspection: response to Malach. Trends Cogn Sci 15:508-509.

21. LeDoux JE (2015) Anxious: Using the brain to understand and treat fear and anxiety. New York: Viking.
22. Carmel D, Raio C, Phelps EA, Carrasco M (2011) Fast unconscious fear conditioning. Journal of vision 11:314-314.
23. Raio CM, Carmel D, Carrasco M, Phelps EA (2012) Nonconscious fear is quickly acquired but swiftly forgotten. Current biology : CB 22:R477-479.
24. Sligte IG, Scholte HS, Lamme VA (2009) V4 activity predicts the strength of visual short-term memory representations. J Neurosci 29:7432-7438.
25. Bergstrom F, Eriksson J (2014) Maintenance of non-consciously presented information engages the prefrontal cortex. Front Hum Neurosci 8:938. PMC4240068.
26. Pan Y, Lin B, Zhao Y, Soto D (2014) Working memory biasing of visual perception without awareness. Atten Percept Psychophys 76:2051-2062.
27. Eriksson J, Vogel EK, Lansner A, Bergstrom F, Nyberg L (2015) Neurocognitive Architecture of Working Memory. Neuron 88:33-46. PMC4605545.
28. Brown R (2014) Consciousness doesn't overflow cognition. Frontiers in psychology 5:1399. PMC4255486.
29. Crick F, Koch C (2003) A framework for consciousness. Nature neuroscience 6:119-126.
30. Jacob J, Jacobs C, Silvanto J (2015) Attention, working memory, and phenomenal experience of WM content: memory levels determined by different types of top-down modulation. Frontiers in psychology 6:1603. PMC4610135.
31. Jacobs C, Silvanto J (2015) How is working memory content consciously experienced? The 'conscious copy' model of WM introspection. Neurosci Biobehav Rev 55:510-519.

BOX 5: RICH VERSUS SPARSE VIEWS ON PHENOMENOLOGY OF PERCEPTION

Our conscious experience of the world appears to be rich and detailed. Yet evidence suggests that large changes in the environment can fail to be detected[1]. So-called change blindness findings suggest that our conscious experience may be much sparser than it appears to be—we may have the feeling that there is a lot of detail represented in our conscious experience, but maybe this is not the case. The question of whether conscious perception is rich or sparse is a heatedly debated topic in consciousness research today. This is an important debate definitive evidence in one direction or the other would have important implications for the broader debate over the nature of consciousness.

Block[2, 3] is a vocal proponent of the rich phenomenology view. His argument roughly takes the following form. In experiments where people are asked to verbally report on what they saw during a brief exposure to a group of stimuli[4-6] they typically say they saw the entire array[2] but can describe only a few of these items. However, if they are cued in a certain way after the stimulus is no longer present they report enough items to suggest that they must have encoded all or most of the relevant information. For example, in Sperling's classic experiments subjects are cued with an audio tone after the array of letters is no longer present. This audio tone cues them to name the letters in either the top (high tone), middle (middle tone) or bottom (low tone) rows. Subjects can get all of most of the letters in the cued rows and since any row could have been cued it is inferred that most of the letters and their specific identities had to be represented prior to the audio cue. According to Block, this suggests that we have rich conscious phenomenology that is distinct from our cognitive access to and ability to report about it. After all, these subjects say that they have a conscious experience of all of the letters and that they use this conscious experience to guide their choices[2]. Block concludes that there are

states which are phenomenally conscious but unaccessed by cognitive systems, and thus unreportable, and he uses this evidence to argue that phenomenal consciousness exists independent of access consciousness, which reflects introspection and cognitive access (this position is discussed further in Box 6).

Block's position has been challenged by researchers who argue for a different interpretation of the data[7-9]. In particular, the opponents suggest that subjects may have sparse conscious representations (as reflected in reports) but detailed unconscious representations (that cannot be reported but that can guide behavior). Both Block's interpretation of the data, and the interpretation proposed by the critics, are consistent with the reports that subjects give of seeing all of the letters, etc. On the sparse phenomenology view (the view of the critics) subjects have a degraded conscious experience of the letters until the cue directs their attention and they then have a detailed conscious experience of the relevant row. Subjects may consciously experience enough of the letters to make it true that they felt like they saw them all. Block's rich phenomenology interpretation, on the other hand, assumes that one has a phenomenally conscious experience of all of the letters in the array but can only access a few of those items. These rich representations, according to Block, are accessed enough to notice that they are letters, and to notice how many letters there are; there is phenomenally conscious detail which could be accessed, but that one cannot in that moment access in such a way as to report their specific identities. Block thus appeals to a hypothetical account of what it is like for the subjects to make his case.

Bronfman et al have recently used findings about color perception to argue a similar point as Block—that color properties in unattended areas of peripheral vision are phenomenally experienced even though unreportable[3, 10]. Specifically, they presented subjects with an array of letters that varied in color diversity from high to low. Subjects were pre-cued (that is, the cue was shown before the onset of the letter array) to attend to a specific row and then afterwards a post-cue indicated which letter from the pre-cued row they should report. Subjects were also asked to estimate the color diversity in unattended rows, which they were able to do, and doing so did not interfere with their ability to recall the post-cued letter. The authors take this as evidence that phenomenal consciousness can be separated from the introspective mechanisms that allow access and enable report. However, as noted above for the Sperling letter cuing task, the data are also consistent with the view that non-conscious information drives the performance of subjects[8, 11, 12].

Thus there is currently insufficient reason to take the experiments discussed in this Box as providing evidence for the claim that unaccessed, unreportable, rich conscious representations exist. Block's key appeal to what it is like for subjects in these experiments is not decisive. The sparse phenomenology position predicts that subjects will experience the array in a partial and degraded way but these generic representations may still label the degraded elements as letters, perhaps even as having determinate detail. We conclude that these data do not allow definitive conclusions relevant to the first-order vs. higher-order debate.

Box 5 References

1. Rensink RA, O'Regan JK (1997) TO SEE OR NOT TO SEE: The Need for Attention to Perceive Changes in Scenes. Psychological Science 8:368-373.
2. Block N (2007) Consciousness, accessibility, and the mesh between psychology and neuroscience. Behav Brain Sci 30:481-499; discussion 499-548.
3. Block N (2014) Rich conscious perception outside focal attention. Trends Cogn Sci 18:445-447.
4. Sperling G (1960) The information available in brief visual presentations. Psychological Monographs 74:1-29.
5. Landman R, Spekreijse H, Lamme VA (2003) Large capacity storage of integrated objects before change blindness. Vision Res 43:149-164.
6. Sligte IG, Scholte HS, Lamme VA (2008) Are there multiple visual short-term memory stores? PLoS One 3:e1699. PMC2246033.
7. Kouider S, de Gardelle V, Sackur J, Dupoux E (2010) How rich is consciousness? The partial awareness hypothesis. Trends Cogn Sci 14:301-307.
8. Brown R (2012) The myth of phenomenological overflow. Conscious Cogn 21:599-604.
9. Phillips IB (2011) Perception and Iconic Memory: What Sperling Doesn't Show. Mind & Language 26:381-411.
10. Bronfman ZZ, Brezis N, Jacobson H, Usher M (2014) We see more than we can report: "cost free" color phenomenality outside focal attention. Psychol Sci 25:1394-1403.
11. Brown R (2014) Consciousness doesn't overflow cognition. Frontiers in psychology 5:1399. PMC4255486.
12. Phillips I (2016) No watershed for overflow: Recent work on the richness of consciousness. Philosophical Psychology 29:236-249.

Box 6: Role of Awareness in First-Order Phenomenal Consciousness

First-order phenomenal consciousness has that awkward property of being a conscious experience that you are not aware of. To get around this problem Block, a leading proponent of first-order theory, has introduced the notion of "awareness access"[1]. In doing so, he acknowledges that some kind of awareness of our first-order states is a necessary component of phenomenal consciousness, but he denies that the relevant kind of awareness is anything like the cognitive kind invoked in higher-order theories.

One suggestion Block makes is that some kind of *deflationary awareness* may suffice. The deflationary notion is inspired by remarks of Ernest Sosa[2] who notes that one smiles one's own smiles but the smiling is not something in addition to the smile. Just having the smile is itself smiling the smile. So, on the deflationary view, saying a conscious state is one we are aware of it is a bit like saying that smiling is always smiling one's own smile. That is, on the deflationary view, we are conscious of our first-order states but not because of any kind of distinct higher-order awareness. To have the state is to be conscious of it, and nothing else is required.

The problem with this view is that it is unable to distinguish conscious states from non-conscious states. In fact, this kind of deflationary awareness seems to accompany every state of the brain, which then, would make all brain states phenomenally conscious (and to make matters worse, if we allow that there are phenomenally conscious states that are not cognitively accessed we might not even be able to tell that these conscious states are there!). As a result, deflationary awareness does not seem to solve the unaware nature of first-order phenomenal consciousness.

A related alternative suggested by Block[1] is that *same-order awareness* might do the job. The traditional distinction between first and higher-order

representational contents is a distinction between what kinds of things are being represented. First-order contents represent states of the environment and higher-order contents represent other mental states. On the traditional higher-order theory these two distinct kinds of representational contents correspond to numerically distinct representational states (at least two, in the simplest case). With same-order awareness, there is one phenomenally conscious state that represents itself[3]. That is, there is said to be one state, a part of which can be understood as having first-order content and another part of the same state with the appropriate higher-order content. Same-order theories thus endorse the view that some kind of inner awareness is required for phenomenal consciousness, like higher-order theories, but deny that this is because there are two states, one of which represents the other.

But what kind of self-representation is at stake here? If the higher-order part of the state turns out to be something cognitive then this is just a variant of the kind of awareness invoked by the higher-order approach and so would not constitute a genuine alternative. Block wants a notion of self-representation that is non-cognitive, and thus not anything like representations invoked by global workspace theories, higher-order theories, or representational theories in general[1]. But what that could possibly be remains unclear at this point. Block has speculated that there may be a notion of self-representation that is robust enough to satisfy the commitment that awareness is crucially bound up in phenomenal consciousness and also the constraint that it be non-cognitive but he doesn't himself offer an account what that would be like.

When pushed, Block[4] has gestured toward what has recently been called the Joint-Determination View[5,6,7]. On this kind of view the first-order and higher-order states jointly determine the nature of one's phenomenal consciousness. If, to illustrate, one were seeing purple, then one would have a first-order state representing purple in the world and a higher-order state representing that the first order state is reliable, something we might express with 'I am vividly perceiving the content of representation F'[6]. By itself the first-order representation of purple results in no conscious experience but when one becomes aware of oneself as being in that state, by having the higher-order decision process 'judge' the first-order state to be reliable, one has a phenomenally conscious purple experience. When one has just the higher-order state the joint-determination view still claims that one will have a phenomenally conscious experience. It will appear to the subject as though they are having a very vivid experience but without any specific quality entering into the experience. Though consistent with the data and not without its advantages, the joint-determination view abandons the explanatory project of the higher-order approach in that it is not able to explain why the conscious purple experience is experienced as purple as opposed to any other quality, or none at all (for further discussion of this issue, see[6,7]).

Part of the strength of the higher-order approach in general is that it appeals to well-defined cognitive processes with known neural underpinnings in the explanation of what inner awareness amounts to. These processes, in turn, are postulated already in theories aimed at explaining mental functioning, and are seemingly needed independently in order to explain how perception works. It is thus natural to extend them to other psychological phenomena like consciousness.

The alternative, perhaps, is a kind of acquaintance. On this view one might postulate the existence of basic qualitative properties, like red, and so on, and then argue that when these qualities are conscious we come to be directly acquainted with them[7]. Though this is a possibility it comes with the heavy price of postulating that mental qualities are basic components of reality, which amounts to something like panpsychism.

Panpsychism has been taken seriously by several authors recently[8,9]. One version of this view holds that the mental qualities, qualia, like red, the sound of a bell, anger, etc. are fundamental elements of reality like mass and charge are. In addition to the fundamental mental qualities one would also need to postulate a primitive notion of awareness that was not explainable in terms of the functioning of the system. While these views are not ruled out a priori we feel they come at too steep a cost; especially when we have more physically amenable candidates that have not been ruled out.

Thus it appears that if we truly want to abandon cognitive access we must also abandon the view that consciousness can be explained at the psychological level. This is a drastic move which is not mandated by the current data. We have cognitive notions of awareness which need to be postulated in order to explain mental functioning. These processes can also be used to explain phenomenal consciousness and unless there is something wrong about the account we ought not to abandon it for speculative metaphysical positions that are more extravagant.

Box 6 References

1. Block N (2007) Consciousness, accessibility, and the mesh between psychology and neuroscience. Behav Brain Sci 30:481-499; discussion 499-548.
2. Sosa E (2002) Privileged Access. In: Consciousness: New philosophical perspectives (Smith, Q. and Jokic, A., eds), pp 273-294 Oxford: Oxford University Press.
3. Kriegel U (2009) Subjective Consciousness: A self-representational theory. Oxford: Oxford University Press.
4. Block N (2011) Response to Rosenthal and Weisberg. Analysis 71:443-448.
5. Lau H (2008) A higher-order Bayesian decision theory of consciousness. In: Models of brain and mind: Physical, Computational, and Psychological Approaches (Banerjee, R. and Chakrabarti, B. K., eds), pp 35-48 Oxford: Elsevier.
6. Brown, R (2012). The Brain and its States. In Shimon Edelman, Tomer Fekete & Neta Zach (eds.), *Being in Time: Dynamical Models of Phenomenal Experience*. John Benjamins. pp. 211-238.
7. Maniscalco B, Lau H (2016) The signal processing architecture underlying subjective reports of sensory awareness. Neurosci Conscious 2016. PMC4972343.
8. Chalmers D (2015) Panpsychism and Panprotopsychism. In: Consciousness in the Physical World: Perspectives on Russellian Monism (Alter, T. and Nagasawa, Y., eds), pp 246-276 Oxford: Oxford University Press.
9. Tononi G, Koch C (2015) Consciousness: here, there and everywhere? Philosophical transactions of the Royal Society of London Series B, Biological sciences 370. PMC4387509.

BOX 7: HIGHER-ORDER THEORIES AND THEIR RELATION TO OTHER COGNITIVE THEORIES OF CONSCIOUSNESS

It is important to make a distinction between the neural notion of higher-order theory, which assumes that lower-order neural representations (say in visual cortex) are not, on their own, equivalent to introspective awareness, from philosophical versions of higher-order theory, such as higher-order thought theory, which specifically require a form of higher-order inner awareness for phenomenal consciousness. Thus, theories based on attention and/or working memory[1-12], or processing by a global workspace[13-17], or interpretation of experience[18, 19] can be construed as higher-order in the neural sense of requiring more than sensory processing in order for consciousness to occur (see **Table 1** for a summary of cognitive theories of consciousness).

For example, Rosenthal argues that global workspace and attentional theories, which call upon higher-order neural areas but not higher-order awareness, are first-order theories[20]. Others sympathetic to Global Workspace theories have argued that the global states themselves amount to a kind of higher-order awareness[21]. In fact Baars himself, the originator of Global Workspace Theory, suggests that the global workspace may only be a necessary condition for consciousness[14]. Sufficiency may require access by an executive self-system, and this sounds a lot like a kind of higher-order awareness. Thus, in this way, every theory of consciousness can be seen as either a higher- or first-order theory. Any theory of consciousness can be interpreted as invoking a kind of higher-order awareness, in which case it is a version of higher-order theory, or it will deny a role for any such awareness, in which case it will be a first-order theory.

Studies like those of Lau and Passingham[22] provide some support for the prediction by the higher-order theorists. It does seem that we can generate cases where task performance is matched, as measured by a measure like d', and yet subjects seem to have differing conscious experiences[22]. If this is so then it seems to count against first-order versions of global workspace. As Lau and Rosenthal[23] note it is possible for a global workspace theorist to posit two distinct channels--one for conscious contents and one for unconscious processes. But computer models of the Lau and Passingham data set suggest that global workspace theories do not capture the known data as well as the models embodying a higher-order structure.[24] The key claim of these kind of dual channel Global Workspace models is that a conscious and an unconscious channel work together to explain both the task performance and the subjective reports while the hierarchical models posit that task performance is mostly driven by first-order states and subjective reports are the result of late stage higher-order processes.

Table 1: Overview Contemporary Theories of Consciousness

First-Order Theory
Premise: the presence of a state is sufficient for the conscious experience of that state. Consciousness consists in outer awareness
Representative Proponents: Ned Block[a], Victor Lamme[b], Fred Dretske[c], Michael Tye[d]

Higher-Order Theory
Premise: the presence of a state is not sufficient for the conscious experience of that state. The state must be re-represented by an additional (higher-order) state. Consciousness consists in inner awareness
Representative Proponents: David Rosenthal[e], Josh Weisberg[f], Richard Brown[g], Hakwan Lau[h]

Global Workspace Theory
Premise: consciousness results from the broadcasting of a state throughout a global processing network
Representative Proponents: Stanislas Dehaene[i], Jean-Pierre Changeux[i], Bernard Baars[j]

Integrated Information Theory
Premise: consciousness is inherent in integrated information
Representative Proponents: Giulio Tononi[k], Christof Koch[k]

Other Cognitive Theories
Premise: consciousness results from cognitive processes such as attention, working memory, metacognition, interpretation of experience
Representative Proponents: Jesse Prinz[l], Daniel Schachter[m], Chris Frith[n], Michael Gazzaniga[o]

Table 1 References

a. Block N (2007) Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav Brain Sci* 30(5-6):481-499; discussion 499-548.
b. Lamme VAF (2005) Independent neural definitions of visual awareness and attention. *Cognitive Penetrability of Perception: Attention, Action, Strategies, and Bottom-Up Constraints*, ed Raftopoulos A (Nova Science Publishers, New York), pp 171-191.
c. Dretske F (1995) Naturalizing the Mind. Cambridge, MA: MIT Press.
d. Tye M (2000) Consciousness, Color, and Content. Cambridge, MA: MIT Press.
e. Rosenthal DM (2005) Consciousness and mind. Oxford: Oxford University Press.
f. Weisberg J (2011) Misrepresenting consciousness. Philosophical Studies 154:409-433.
g. Brown R (2012) The myth of phenomenological overflow. Consciousness and cognition 21:599-604.
h. Lau H, Rosenthal D (2011) The higher-order view does not require consciously self-directed introspection: response to Malach. Trends Cogn Sci 15:508-509.
i. Dehaene S, Changeux JP (2011) Experimental and theoretical approaches to conscious processing. Neuron 70:200-227.
j. Baars BJ (2005) Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. Prog Brain Res 150:45-53.
k. Tononi G, Boly M, Massimini M, Koch C (2016) Integrated information theory: from consciousness to its physical substrate. Nat Rev Neurosci 17:450-461.
l. Prinz JJ (2012) The Conscious Brain: How Attention Engenders Experience. New York: Oxford University Press.
m. Schacter DL (1998) Memory and awareness. Science 280:59-60.
n. Frith C, Perry R, Lumer E (1999) The neural correlates of conscious experience: an experimental framework. Trends in Cognitive Sciences 3:105-114.
o. Gazzaniga MS (2012) Who's in Charge?: Free Will and the Science of the Brain. New York: Ecco.

Table 2: Some Variants of Higher-Order Theory

Traditional Higher-Order Thought (HOT) Theory
Premise: a thought-like mental state makes one aware of oneself as being in a first-order state, by representing oneself as being in that state, thereby making the first order state conscious.
Role of Empty Higher-Order States? Higher-order states refer to non-existent first-order states, which can be thought of as the conscious state (the state one represents oneself as being in).
Representative Proponents: David Rosenthal[aa], Josh Weisberg[bb]

Higher-Order Representation of a Representation (HOROR) Theory
Premise: a thought-like mental state makes one aware of oneself as being in a first-order state, by representing oneself as being in that state, and this constitutes phenomenal consciousness.
Role of Empty Higher-Order States? Appropriate higher-order states are phenomenally conscious.
Representative Proponents: Richard Brown[cc], Joseph LeDoux[dd]

Same-Order (SO) Theory
Premise: a mental state referring to, or representing, itself, is a conscious state. This complex state, consisting of a world directed component and a self-referential component, constitutes the subject being aware of the conscious state.
Empty-Higher-Order Thoughts? These are irrelevant. When the first-order component is missing there is no resulting conscious state.
Representative Proponents: Uriah Kriegel[ee], Rocco Gennaro[ff], Robert VanGulick[gg]

Joint-Determination (JD) Theory
Premise: A higher-order cognitive decision process whereby an incoming first-order perceptual signal is 'judged' to be reliable (or not) is a conscious state. The resulting phenomenal consciousness depends jointly on the nature of the first-order signal and the 'judgement' by the higher-order process.
Empty Higher-Order States? When first-order component is missing there is still conscious experience associated with the higher-order element alone, but this conscious experience is partial or degraded without the first-order element.
Representative Proponents: Hakwan Lau[hh]

Table 2 References

aa. Rosenthal DM (2005) Consciousness and mind. Oxford: Oxford University Press.
bb. Weisberg J (2011) Misrepresenting consciousness. Philosophical Studies 154:409-433.
cc. Brown R (2015) The HOROR theory of phenomenal consciousness. Philosophical Studies 172:1783-1794.
dd. LeDoux JE (2015) Anxious: Using the brain to understand and treat fear and anxiety. New York: Viking.
ee. Kriegel U (2012) Precis of subjective consciousness: a self-representational theory. PhilosophicalStudies 159:443-445.
ff. Gennaro RJ (2011) The Consciousness Paradox. Cambridge, MA: MIT Press.
gg. Van Gulick R (2004) Higher-order global states (HOGS): An alternative higher-order model of consciousness. In: Higher-Order Theories of Consciousness: An Anthology (Gennaro, R. J., ed): John Benjamins.
hh. Maniscalco B, Lau H (2016) The signal processing architecture underlying subjective reports of sensory awareness. Neurosci Conscious 2016. PMC4972343.

Box 7 References

1. Shallice T (1988) Information processing models of consciousness. In: Consciousness in contemporary science (Marcel, A. and Bisiach, E., eds), pp 305-333 Oxford: Oxford U. Press
2. Baddeley A (2000) The episodic buffer: a new component of working memory? Trends Cogn Sci 4:417-423.

3. Schacter DL (1989) On the relation between memory and consciousness: Dissociable interactions and conscious experience. In: Varieties of memory and consciousness: Essays in honour of Endel Tulving (Roediger, H. L. I. and Craik, F. I. M., eds), pp 355-389 Hillsdale,NJ: Lawrence Erlbaum Associates.
4. Schacter DL, Buckner RL, Koutstaal W (1998) Memory, consciousness and neuroimaging. Philosophical transactions of the Royal Society of London Series B, Biological sciences 353:1861-1878.
5. Frith CD (2008) The social functions of consciousness. In: Frontiers of Consciousness: Chichele lectures (Weiskrantz, L. and Davies, M., eds), pp 225-244 Oxford: Oxford University Press.
6. Prinz JJ (2012) The Conscious Brain: How Attention Engenders Experience. New York: Oxford University Press.
7. Crick F, Koch C (2003) A framework for consciousness. Nature neuroscience 6:119-126.
8. Jacobs C, Silvanto J (2015) How is working memory content consciously experienced? The 'conscious copy' model of WM introspection. Neurosci Biobehav Rev 55:510-519.
9. Johnson-Laird PN (1988) The computer and the mind: An introduction to cognitive science. Cambridge, Mass.: Harvard University Press.
10. Cohen MA, Cavanagh P, Chun MM, Nakayama K (2012) The attentional requirements of consciousness. Trends Cogn Sci 16:411-417.
11. Bor D, Seth AK (2012) Consciousness and the prefrontal parietal network: insights from attention, working memory, and chunking. Frontiers in psychology 3:63. PMC3298966.
12. Baars BJ, Franklin S (2003) How conscious experience and working memory interact. Trends Cogn Sci 7:166-172.
13. Baars BJ (1988) A Cognitive Theory of Consciousness. New York: Cambridge University Press.
14. Baars BJ (2005) Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. Progress in brain research 150:45-53.
15. McGovern K, Baars BJ (2007) Cognitive Theories of Consciousness. In: The Cambridge Handbook of Consciousness (Zelazo, P. D. et al., eds), pp 177-205 New York: Cambridge University Press.
16. Dehaene S, Changeux JP (2011) Experimental and theoretical approaches to conscious processing. Neuron 70:200-227.
17. Dehaene S, Naccache L (2001) Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. Cognition 79:1-37.
18. Gazzaniga MS (2008) Human: The science behind what makes us unique. New York: Ecco.
19. Gazzaniga MS (2015) Tales from both sides of the brain: A life in neuroscience. New York: Ecco/Harper Collins.
20. Rosenthal D (2012) Higher-order awareness, misrepresentation and function. Philosophical transactions of the Royal Society of London Series B, Biological sciences 367:1424-1438.
21. Van Gulick R (2004) Higher-order global states (HOGS): An alternative higher-order model of consciousness. In: Higher-Order Theories of Consciousness: An Anthology (Gennaro, R. J., ed): John Benjamins.
22. Lau HC, Passingham RE (2006) Relative blindsight in normal observers and the neural correlate of visual consciousness. Proceedings of the National Academy of Sciences of the United States of America 103:18763-18768. PMC1693736.
23. Lau H, Rosenthal D (2011) Empirical support for higher-order theories of conscious awareness. Trends Cogn Sci 15:365-373.
24. Maniscalco B, Lau H (2016) The signal processing architecture underlying subjective reports of sensory awareness. Neurosci Conscious 2016. PMC4972343.

BOX 8: PHENOMENAL CONSCIOUSNESS IN HIGHER-ORDER THOUGHT THEORIES OF CONSCIOUSNESS

Some higher-order theorists, like David Rosenthal, construe 'phenomenal consciousness' as denoting a kind of consciousness that sensory qualities have independently of any kind of cognitive awareness—the properties by which we represent physical features like the shape or color of visual stimuli or the pitch of

sounds[1]. Higher-order representation then makes the first-order state conscious (see the arrow from working memory networks to visual cortex in Figure 2b). Rosenthal takes the debate between himself and first-order theorists like Block to be about whether these sensory qualities can occur independent of consciousness. As Rosenthal uses the terms, there is no phenomenal consciousness, since the sensory properties only become conscious via higher-order states. Construed this way the existence of phenomenal consciousness would count against any kind of higher-order theory. Rosenthal, for example, argues that pain can occur unconsciously, and when it does it has the same qualitative character (i.e. sensory properties) as it does when it is conscious, except it is not like anything for the subject. He means by this that there is no kind of consciousness that these unconscious pains have. Rosenthal is obviously not a first-order theorist, but rather he interprets 'phenomenal consciousness' to be a posit of first-order theories. However, in the more general, and less restricted, sense of phenomenal consciousness that we employ here, phenomenal consciousness is the common explanatory target of both higher-order and first-order theories. Use of this theory-neutral sense of phenomenal consciousness makes phenomenal consciousness what higher-order theory seeks to account for.

Box 8 References

1. Rosenthal D (2011) Exaggerated reports: reply to Block. Analysis 71:431-437.

BOX 9: RELATIONAL VERSUS NON-RELATIONAL HIGHER-ORDER THEORIES

The possibility of a mismatch between the content of first and higher-order states has long been thought to pose a problem for higher-order theories. For instance, if one has a lower-order state representing that one is seeing a green leaf and yet, for whatever reason, comes to have a higher-order state representing that one is seeing a red leaf what is it like for the subject? Higher-order theorists have for the most part responded that this is not a theoretical possibility. However, empirical reasons suggest that this actually happens and that it does so in a way that favors a particular kind of higher-order thought theory.

For instance, Brown[1] argues that we have cases of mismatch resulting from inattentional inflation and a rare form of Charles Bonnett syndrome. In some rare forms of Charles Bonnett syndrome there is extensive damage to visual cortex and yet subject's report vivid hallucinations. Even if we assume that the residual cortex is sufficient for some first-order representations to survive it is reasonable to think that this would not be enough to support the rich and detailed conscious experience these subjects report. Thus this looks like a case where there is more in the conscious experience of the subject than can be accounted for by the first-order states. In inattentional inflation subjects seemingly overestimate their visual phenomenology in unattended areas. They may report high confidence in seeing something in an unattended region even though there may have been nothing to see (and vice versa) and when their performance on unattended stimuli is not better than to attended stimuli.

In a more commonsense vein, David Rosenthal[2] has argued that this is a common occurrence in ordinary conscious experience. He has suggested that it is often the case that when we consciously see something as red we may not see it as an exact shade of red, though presumably we represent it as some specific shade of red at that lower-order level. If these views are correct then mismatch between levels cannot be a problem for higher-order theories; in fact these cases suggest that the mismatch between levels gives rise to a prediction which is empirically supported.

But what about the arguably more extreme case where the first-order state is altogether absent? In these 'empty' higher-order thought cases the state which the higher-order state represents is missing and some, notably Ned Block, have argued that these kinds of cases pose a serious challenge to higher-order theories. The basic idea of his challenge can be brought out by asking which state it is that is phenomenally conscious? The first-order state does not exist (by stipulation) and so it cannot be phenomenally conscious. The response by the HOROR theory is to argue that this shows that it is the higher-order state which is phenomenally conscious. Rosenthal has suggested that in these kinds of cases it is the notional non-existent first-order state which is phenomenally conscious. This sounds very odd to some readers but all that he means by it is that the state which is conscious is the state that the higher-order thought attributes to one. While this may sound paradoxical – a conscious state that doesn't exist! - all that it means is that the state one seems to be in does not actually occur. While this may be correct there is still a strong intuitive pull to the claim that phenomenally conscious states must exist. HOROR theory is a useful alternative to the traditional approach and answer that Rosenthal defends. On the HOROR theory it is the higher-order state which is phenomenally conscious. Not because it is represented by some further higher-order state, though that may occur, but because it is the kind of state that allows one to be aware of one's own mental life, which, the theory claims, is all there is to phenomenal consciousness.

Box 9 References

1. Brown, Richard (2012). The Brain and its States. In Shimon Edelman, Tomer Fekete & Neta Zach (eds.), *Being in Time: Dynamical Models of Phenomenal Experience*. John Benjamins. pp. 211-238.
2. Rosenthal D (2011) Exaggerated reports: reply to Block. Analysis 71:431-437.

BOX 10: RESEARCH QUESTIONS

1. Damage to first-order subcortical circuits (including defensive survival circuits, such as those involving the amygdala) disrupts the expression of objective (behavioral and physiological) responses elicited by threats. The theory proposed here suggests that people with such lesions should still experience fearful feelings, but that the feelings might be muted. Some evidence, discussed in the paper, supports the conclusion that fearful feelings can persist in patients with amygdala damage, but this should be studied further. Whether quantitative changes in fearful experiences occur with amygdala lesions has not been studied. If such changes occur, it would be

important to determine whether they are due to alterations of signals within the brain itself or of feedback signals from the body, or both. Quantitative differences may be especially important in intense, truly fearful, experiences where brain arousal and body feedback may be especially prominent. However, for ethical reasons, experimental subjects cannot be placed in situations that arouse intense fearful or other emotional experiences. This poses limits on the ability to fully address such questions with standard methods and research practices.

2. Activity in prefrontal cortex is correlated with introspective awareness of threats. We propose that the same circuitry underlies the experience of "fear" elicited by threats. Evidence involving studies of visual masking and blindsight patients discussed in the paper suggests that this is the case, but relatively few studies have obtained reports about feelings in studies involving masked threats or blindsight.

3. Our theory assumes a representational-hierarchical relation between first-order subcortical circuits (e.g. amygdala) and higher-order cortical circuits (e.g. prefrontal) in the experience of fear. However, unlike in perception, for emotional experiences, the lower-order (amygdala) and higher-order (prefrontal) circuits are activated in parallel by sensory inputs, and the lower-order state is not necessary for the higher-order state (see above). The amygdala states in our theory are, in fact, one of several factors that that can bias pattern completion of an emotion schema and give rise to the higher-order state and the experience of fear. The contribution of emotion schema to emotional experience is relatively unexplored. An important question is whether and if so how pattern completion leads to a particular emotional experience when stimuli incompletely activate an emotion schema, such as a fear schema. Is the presence of a threat, identified via memory, sufficient to give rise to fear, or is amygdala-triggered brain arousal, or feedback from amygdala-dependent body responses also necessary? And if not necessary, what do they contribute to the experience, if anything?

4. Nonconscious working memory is often studied using brief stimulus exposures, as in masking. It is unclear the extent to which limitations reported in some studies are due to the use of suboptimal viewing conditions in studies of nonconscious processing. Nonconscious conditions involve brief and/or masked exposures or stimuli that compete between the eyes, while conscious processing uses conditions that allow unimpaired, richer processing. In blindsight studies nonconscious processing can be studied using less restricted stimulus exposure conditions. Such studies might be more suitable for assessing the capacity of nonconscious working memory.

5. Most work on the brain basis of consciousness has involved correlations between brain activity and measures of consciousness. For example, activity in lateral and medial prefrontal cortical areas is often correlated with

awareness. Recent work has attempted to go beyond simple correlation and "decode" experience from multivariate pattern analysis of brain activity. As of now the results are mixed. Is this due to the lack of involvement of these areas in the experience or to the fact that existing methods allow us to identify circuits that are necessary for conscious experience but do not allow us to decode the content of the experience? Correlations between reports of awareness and neural activity in specific areas are useful but do not demonstrate the necessity of the activity for the experience. Causal implication of areas requires conditions that interfere with consciousness. While studies of people with brain damage are the traditional way to implicate brain areas in brain function, lesions do not respect anatomical boundaries. Also, because they are permanent, lesions can lead to compensatory changes in undamaged tissue. Some success has been had in disrupting reports of introspective awareness using transcranial magnetic stimulation. Systematic studies in which the effects of stimulation of each of the lateral and medial prefrontal regions that have been implicated through correlations would be useful. Because the key areas are widely distributed in lateral and medial prefrontal cortex, negative results with localized stimulation might need to be followed up with broader coverage. At the same time, improved methods for studying correlations might more accurately pinpoint which regions to target for specific tasks.

6. Higher-order theories make the prediction that one should be able to vary the conscious experience of a subject by manipulating higher-order states and maintaining first-order states. Conversely they also predict that one should be able to hold the conscious experience of subjects constant while varying first-order states. Some evidence suggests this is the case but studies directly addressing this issue would be helpful.

7. To fully empirically differentiate between higher-order and global workspace theories of consciousness (as well as between the variants of these theories) we need to know more about the neural underpinnings of cognition. In particular, it is necessary to be able to distinguish between the neural activity associated with the global neuronal workspace from activity associated with the kind of higher-order awareness postulated by higher-order theories.

8. Future research will need assess the possible contribution of different lateral and medial prefrontal circuits to determine which circuits and cognitive processes contribute to nonconscious representations as opposed to possible higher-order representations of these that render the processing conscious. In addition, the question of whether different aspects of both conscious and nonconscious processing involve different subcircuits is important.

9. Discussions of consciousness tend to emphasize prefrontal circuits. Other areas, especially parietal and insular areas, are mentioned in passing but are not always given as much consideration. Future work should evaluate these

areas side by side with prefrontal areas in an effort to more accurately identify which circuits contribute to which cognitive processes underlying specific aspects of conscious experience.

10. The role of the self in consciousness, including emotional consciousness, is relatively unexplored, as is the neural processes that underlie self-representations, including self-schema. This is a ripe area for research.