

Winner of the 2015 Tony Chadwick Essay Prize

Are Beliefs “In the Head”?

by John Harries

This essay addresses the location of beliefs and suggests that they may not be located solely in the head or indeed be determined solely by the brain. Such a possibility is of course in contradiction to the dominant scientific view espoused by Hippocrates (460 to 370 BC) and more recently by the neurologist Dick Swaab that “*everything we think, do and refrain from doing is determined by the brain*” (Swaab 2014). In this essay I draw on the work of a number of philosophers, psychologists and cognitive archaeologists to suggest that this common sense idea may not be completely correct.

In his influential essay “The meaning of “meaning”” Hilary Putnam (1975) concluded that “*“meanings” just ain’t in the head*”. If this is true then beliefs, which are expressed with words whose meanings are not in the head, are themselves not in the head. This view on the location of belief is one aspect of a philosophical position called Externalism, specifically “semantic externalism”.

Putnam reaches this radical conclusion via a series of thought experiments the most celebrated of which involves Earth, its replica Twin-Earth and their inhabitants, respectively Oscar and his doppelgänger Twin-Oscar. The only difference between the planets is that on Earth water has the chemical constitution H₂O whilst on Twin-Earth the life sustaining fluid, which in all phenomenological respects resembles water and is called water by Twin-Oscar, has the formula XYZ. Putnam asserts that when Oscar travels to Twin-Earth and he and Twin-Oscar refer to the clear fluid in rivers etc. as “water” Twin-Oscar is speaking truthfully because “water” designates XYZ on Twin-Earth but Oscar is speaking falsely because for him “water” designates H₂O. Furthermore this would have been true before 1750 when chemical discoveries of molecular composition were made and any inhabitant of Earth or Twin-Earth could have been aware of the essence (and the differences) of what they both called “water”. On Putnam’s account, if I (an Earth dwelling English speaker) am thinking of water I am thinking of H₂O even if I don’t know it.

This divergence of meaning of the word water on the two planets results from a rigid designation of the word water (and other natural-kind terms). On Earth the essence of the substance to which the word refers is H₂O and on Twin-Earth it is XYZ. If you are from Earth and you believe that XYZ is water you are making a mistake and your belief is false. This is because, in Putnam’s formulation meaning (and therefor beliefs) are an aspect of knowledge and are factive. If Oscar incorrectly looks at XYZ and says “that is water” the utterance does not constitute a belief.

The importance of this thought experiment is that Oscar and Twin-Oscar are identical and their narrow psychological states (those states inside their head and independent of the external world) are identical when they are looking at H₂O and XYZ respectively. Their functional, brain and phenomenal states are all the same but they mean different things when they look at the fluid and say water. Thus meaning “*ain’t in the head*”. McGinn (1977) builds on this thought experiment by imagining that Oscar on Earth and Twin-Oscar on Twin-Earth both utter the belief that “water quenches thirst”. Oscar is referring to H₂O and Twin-Oscar to XYZ. The two Oscars have identical intrinsic properties but hold beliefs with different truth conditions. For McGinn this demonstrates that “*some beliefs do not supervene on intrinsic facts and therefor Externalism is true*” (Stanford Encyclopedia of Philosophy 2014)

Putnam generalises from this fictional experiment to a personal real life example. He says that he cannot identify Beech or Elm trees or tell the difference between them. However when he uses the term Beech he means the same extension of all the Beech trees as when any other English speaker uses the word. Similarly when he uses the word Elm he means the same set of trees as anyone else referring to Elm. However he can't differentiate between Elm and Beech so the difference in meaning is not inside his head. There is here a "*division of linguistic labour*" between people who use the words without really understanding their meaning (such as Putnam) and experts who could identify the trees and really know what the words mean. "*not everyone to whom the distinction is important needs to be able to make the distinction*". Putnam is using words whose meaning resides outside his head and in the heads of (in this example) tree experts.

In a similar thought experiment Burge (1979) imagines an individual with rheumatoid pains in this world and in a counterfactual twin world. In this world, suffering a pain in his thigh he goes to the doctor complaining of arthritis. The doctor informs him that as arthritis is an inflammation of joints he must be incorrect in his self diagnosis. In the counterfactual twin world the exact same individual goes to the doctor asserting the same belief but in the twin world the linguistic norms are different and the word arthritis includes pains of the bones. In twin world the patient's declaration that he has arthritis (in our world we would have to give this rheumatoid syndrome another name, say "Tharthrititis") is correct. Every internal aspect of the patient in the two worlds remains constant, the only variable is the linguistic norms of the community in this and the counterfactual world. As Kim (2011) observes: "*if this is right, beliefs and other intentional states do not supervene on the internal physical-psychological states of persons; if supervenience is wanted, we must include in the supervenience base the linguistic practices of the community to which people belong*". Again, belief is not purely "in the head" but determined in this case by the social/linguistic environment. Burge's variant of externalism is described as "Social Externalism"

Clarke and Chalmers (1998) propose a third kind of externalism which they dub "*active externalism*", "*based on the active role of the environment in driving cognitive processes*". In their paper "The Extended Mind" they provide a number of examples of cognitive process involving human-computer interaction and a category of behaviour termed "epistemic action" (Kirsch & Maglio 1994) which manipulates the world in order to support cognitive processes such as recognition and search. They draw the boundary of the cognitive system outside the brain and body and into the world because they argue that "*if...a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognising as part of the cognitive process, then that part of the world is part of the cognitive process*". This position of equivalence between intracranial and external world cognitive resources is called the "parity principle".

Clarke and Chalmers extend this reasoning to belief and give an example of an Alzheimer's patient whose beliefs (such as the location of a museum) reside in the notebook which he habitually uses to record information, rather than his brain's unreliable memory system. This repository of beliefs is available in real-time and directly affects behavior. A more contemporary example of what Wilson (2010) has described as "cyborg fantasy" arguments for the extended mind is the Juliana Moore character in "Still Alice" (Sony Pictures 2014) whose early onset Alzheimer's is initially disguised by her extensive use of her smart phones organizational software. Clarke and Chalmers also suggest that cognition and beliefs may be socially extended to relevant others. A close couple for example may "fill in" each others beliefs in the same way as the patient's note book/mobile. Both the device and the partner function as mental prosthesis (analogous to

an artificial limb) the essential attributes of which are “*a high degree of trust, reliance and accessibility*”.

In all of these examples of externalism, language has a critical role in positioning belief outside the head. Clarke and Chalmers echo Putnam when they say that “*the advent of language has allowed us to spread this (cognitive) burden into the world. Language thus construed is not a mirror of our inner states but a complement to them. It serves as a tool whose role is to extend cognition in ways that on-board devices cannot*”. The anthropologist Roger Bartra (2014) stretches this idea still further with his suggestion of an “exocerebrum”, a culturally transmitted “symbolic substitution system” which functions as a mental prosthesis.

As outlined by Wheeler (2010) the parity principle described above has strong philosophical support from functionalism and its concept of “multiple realisability”. In functionalism a mental state “counts as the mental state it does because of the causal relation it bears to sensory inputs, behavioural outputs and other mental states” (Putnam 1967). Multiple realisability breaks from a dependence on neuronal activity to support mental states and is open to this function being performed by any material mechanism including those which extend beyond the body.

The extended mind thesis is not confined to linguistic and other symbolic systems. Wilson (2010) expresses the thesis as follows: “*for at least a variety of cognitive activities, the physical configuration of the brain is not metaphysically sufficient for their performance qua cognitive activities. Something more is needed and that something involves the physical configuration of the world beyond the head*”. Wilson provides examples of the role of physical interaction with things in the act of problem solving the most prosaic of which is completing jigsaws in which the manual manipulation of the pieces is a key aspect of a successful process. He regards this dynamic as a causal integration of things with “onboard” capacities which organisms already have. An interesting example of this phenomena which synthesizes symbolic content and temporospatial manipulation is Hutchins (1988) description of air traffic controllers use of paper strips to represent aircraft and the difficulties encountered in attempting to computerise this process. The strips are coded symbols for planes but the dynamics of multiple contending aircraft are understood through the physical manipulation of the paper strips, the bodily orientation of the controllers and their shared attention to and dialogue regarding these strips. The controllers belief that their planes are on safe trajectories is provisional and dynamic and resides in the interaction of brain, body, movement, external symbolic manipulation and dialogue.

In “A problem for externalism” Marianne Talbot (2014) says “*If mental states are not narrow but wide psychological states, then how are they causally implicated in the production of our behaviour?*” “*Surely one thing of which we can be certain is that our behaviours are a function of what goes on inside our heads.*” I believe that there is reason to doubt that this last assertion is completely correct. At a relatively trivial level there are a range of behaviours which appear to be reflex, such as hand withdrawal from heat, and although neurological they bypass the head. More significantly, in philosophy and the social sciences a “material turn” has occurred which takes the idea of decentralised agency seriously. A founder of this movement is Bruno Latour with Actor-Network Theory (ANT) which conceptualises agency as distributed and possessed, not solely in the heads of human agents but in relational networks of persons and things. As Malafouris (2013) summarises “*Power, intentionality and agency are not properties of the isolated person or the isolated thing; they are properties of a chain of association*”.

Analogous to cognition and the extended mind theory discussed above ANT redraws the systems boundary of behaviours in the world to give equality of agency to things as well as people.

Malafouris (2013) discusses Searle's (1983) philosophy of agency and behaviour which distinguishes between "prior intention" and "intention in action". "Prior intention" is an internal representational state with no pragmatic effect on the world. Prior intention may or may not precede "intention in action" in which the internal intentional state and external movement become inseparable and the "*boundary between the mental and physical collapses.*" According to Malafouris intention in action is "*not an internal property but a component of extended cognition*" and being constituted of both persons and things is not a basis on which to attribute exclusively human agency to material engagement. From the perspective of a cognitive archaeologist Malafouris thesis of material agency is that agency and intentionality is an emergent property of material engagement involving both humans and things.

Varela, Thompson, and Rosch (1991) building on the phenomenology of Merleau Ponty and Husserl and the American Pragmatism of James and Dewey developed the ideas of the embodied mind in which cognition, experience and meaning making is not confined to computational representation in the head but is enacted through "mutual interactions between the physiology of the organism, its sensorimotor circuit and the environment" (Stanford Encyclopedia of Philosophy 2011) the structural coupling of brain body and world. Enactivism and embodiment are closely related. Indeed Varela et al (1991) despite their focus on enactivism entitled their book "The Embodied Mind". Hutchins has described enaction as follows: "*enaction is the idea that organisms create their own experience through their actions. Organisms are not passive receivers of input from the environment, but are actors in the environment such that what they experience is shaped by how they act.*" (Hutchins 1996). I would add that how they act is shaped by what they interact with or indeed what interacts with them, that is by material engagement. This is not a novel idea. Pitt-Rivers (1875 in Hodder 2012) describing the relationship between flints and the prehistoric axes fashioned from them wrote "*so completely does the fabricator appear to have been controlled by the necessities of his art that in tracing these successive forms one is almost tempted to ask whether the principle of causation lay mostly in the flint or the flint worker*".

Bodies, in the sense of perceptual and motor neuron systems and voluntary muscles are essential to enaction but there is more to bodies than these subsystems and there is more to mental life and meaning making than cognition narrowly defined by abstract "problem solving". Bodies also consist of complex cardiovascular and endocrine systems which contribute to the affective life of humans and affect (emotions and mood states etc.) and cognition are intimately linked. Even the archetypal cognitive tasks of problem solving and decision making involve a significant role for affect (see e.g. case studies reported by Damasio 1994). Neurological studies are also identifying a cognitive role for structures previously associated purely with affect e.g. the role of the amygdala in in the cognitive functions of attention and associative learning (Pessoa 2008 in Colombetti 2014).

Both functionally and structurally cognition and affect are entangled. Views on the relative contribution of cognition, body and world to the experience of emotion have shifted back and forth since William James (1884) account of emotion in terms of perception of bodily changes. This argument has gone full circle from the pragmatists via Arnold (1960) and the later cognitivists, who minimised the role of the body in emotion, to enactivists such as Colombetti (2014) who concurs with Dewey's view that the cognitive and emotional

response to the world are not separable except on subsequent reflection. On confronting a bear “the frightful object and the emotion of fear are two words for the same experience” (Dewey 1895). The object, the perception, the emotion, the belief and indeed the action are one. Physiology and the autonomic nervous system, voluntary action and the environment shape belief as much as cognition.

In this essay I have tried to demonstrate a number of ways in which the boundaries of the system which sustains beliefs can plausibly be thought of as extending beyond the believer’s head to include their bodies and the wider world. In summary:

- Beliefs, in the sense of propositional attitudes, are expressed in language. Language is composed of words which have meaning. The arguments of semantic and social externalism demonstrate that beliefs are embodied in and constrained by the designation of words and the linguistic practices of the believer’s community, residing not just in the head but in the world.
- By way of the “parity principle” and the concept of “multiple realisability” Active Externalism extends the boundary of the cognitive system supporting beliefs to any trusted, reliable and accessible prosthetic entity outside the head and in the world.
- Certain types of belief such as “this jigsaw piece will fit here” or “these planes are safe and will not collide” are provisional and dynamic. In such cases the beliefs are constituted by the active manipulation of external objects by the individual. Again, such external objects constitute cognitive prostheses which extend the boundaries of the belief system.
- Concepts of enactivism and material agency decenter the brain. In these views beliefs arise from and reside in the interaction of brain, body and world.
- Affective states shape belief and are determined by the complex interaction of environment, autonomic nervous and endocrine systems. The belief that I see a bear and am in danger resides as much in the environment, my cardiovascular system, my adrenal glands, my skin and my legs as it does in my head.

All of the above leaves me with a question. Do I believe that beliefs are not just in the head? I think the answer must be yes because at present the best articulation of that belief is this paper.

References

- Arnold, M.B. 1960 *Emotion and Personality, Vol. 1: Psychological Aspects* Columbia University Press
- Bartra, R. 2014 *Anthropology of the Brain: Consciousness, Culture and Free Will* Cambridge University Press
- Burge, T. 1979 *Individualism and the Mental* in Chalmers, D.J. ed. 2002 *Philosophy of Mind: Classical and Contemporary Readings* Oxford University Press
- Clarke, A. & Chalmers, D.C. 1998 *The Extended Mind* in Chalmers, D.J. ed. 2002 *Philosophy of Mind: Classical and Contemporary Readings* Oxford University Press
- Colombetti, G. 2014 *The Feeling Body - Affective science meets the enactive mind*. The MIT Press
- Damasio, A. 1995 *Descarte's Error*. Picador
- Dewey, J. 1895 The Theory of Emotion. (II)The Significance of Emotion. *Psychological Review* 2:13-32
- Hodder, I. 2012 *Entangled, an archeology of the relationship between humans and things*. Wiley-Blackwell
- Hutchins, E. 1996. *Cognition in the Wild*. MIT Press.
- Hutchins, E. & Norman, D. A. (1988). Distributed cognition in aviation: a concept paper for NASA (Contract No. NCC 2-591).Department of Cognitive Science. University of California, San Diego.
- James, W. 1884 What is an emotion? *Mind* 9: 188-205
- Kim, J. 2011 *Philosophy of Mind - 3rd Edition* p. 245 Westview Press
- Kirsh, D and Maglio, P. 1994 *On Distinguishing Epistemic from Pragmatic Action*. Cognitive Science 18: 513-49
- Latour, B. 2005 *Reassembling the Social: An introduction to Actor Network Theory*. Oxford University Press
- Lau, Joe and Deutsch, Max, "Externalism About Mental Content", *The Stanford Encyclopedia of Philosophy* (Summer 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2014/entries/content-externalism/>.
- Malfouris, L. 2013 *How Things Shape the Mind: A theory of Material Engagement*. MIT Press
- McGinn, Colin, 1977. "Charity, interpretation, and belief," *Journal of Philosophy*, 74: 521–535.
- Pessoa, L. 2008. On the relationship between emotion and cognition. *Nature Reviews Neuroscience* 9:148-158

Pitt-Rivers, A. Lane-Fox 1874 & 1875 *The Evolution of Culture and Other Essays* Oxford: Clarendon Press.

Putnam, H. 1967. Psychological Predicates. In W.H. Capitan & D.D. Merrill (eds), *Art Mind and Religion*. University of Pittsburg Press.

Putnam, H. 1975 *The Meaning of "Meaning"* in Chalmers, D.J. ed. 2002 *Philosophy of Mind: Classical and Contemporary Readings* Oxford University Press

Searle, J.R. 1983 *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press

Stanford Encyclopedia of Philosophy - Externalism About Mental Content
First published Mon Oct 21, 2002; substantive revision Wed Jan 22, 2014

Swaab, D.F. 2014 *We are our brains*. Allen Lane

Talbot, M. Philosophy of Mind course notes 2014. Oxford University Department of Continuing Education.

Varela, F.J. Thompson, E. Rosch , E. (1992). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press

Wheeler, M. 2010 *In Defense of Extended Functionalism*. In Menary R. (ed) *The Extended Mind*. MIT Press

Wilson, R.A. 2010 *Meaning Making and the Mind of the Externalist*. In Menary R. (ed) *The Extended Mind*. MIT Press