Cain, M

Essentialism, externalism, and human nature

Royal Institute of Philosophy Supplement, 2012, 70 (July). pp. 29-51.
10.1017/S1358246112000033

Essentialism, Externalism and Human Nature

M.J. Cain

1. Introduction

Psychological essentialism is a prominent view within contemporary developmental psychology and cognitive science according to which children have an innate commitment to essentialism. If this view is correct then a commitment to essentialism is an important aspect of human nature rather than a culturally specific commitment peculiar to those who have received a specific philosophical or scientific education.[1] In this article my concern is to explore the philosophical significance of psychological essentialism with respect to the relationship between the content of our concepts and thoughts and the nature of the extra-cranial world. I will argue that, despite first appearances, psychological essentialism undermines a form of externalism that has become commonplace in the philosophy of mind and language.

2. Psychological Essentialism

As its name suggests, psychological essentialism is related to the traditional philosophical doctrine of essentialism.[2] One can draw a rough distinction between two versions of essentialism. According to the first, many of the individual things that

[1] This latter view of essentialism is endorsed by Jerry Fodor, *Concepts: Where Cognitive Science Went Wrong* (Oxford: Oxford University Press, 1998).

[2] Historical advocates of essentialism include Aristotle and Locke. Perhaps the most prominent recent champions of essentialism are Kripke, *Naming and Necessity* (Oxford: Blackwell, 1980) and Putnam, 'The meaning of "meaning"', in his *Mind, Language and Reality: Philosophical Papers Volume 2* (Cambridge: Cambridge University Press, 1975).

populate the world have essences, where an essence is a property (or collection of properties) that is central to the identity of that thing so that it couldn't lose the property without ceasing to exist. For example, it might be claimed that it is part of my essence that I am human but not part of my essence that I am an academic philosopher. Call this essentialism with respect to individuals. According to the second version of essentialism, it is categories of things that have essences. For example, in order to belong to the category HUMAN it is essential to be a mammal. This leaves it open as to whether any individual human is essentially human or as to whether any particular thing has an essence as such (as opposed to an essence relative to a particular category to which they belong). Call this doctrine essentialism with respect to categories.[3]

Psychological essentialism is a view within developmental psychology – and cognitive science more widely – that has come to prominence over the last two decades. In its boldest form it is the view that children are innately essentialist with respect to many of the categories for which they have concepts. For example, in virtue of an innate commitment to essentialism, a child who has acquired the concept DOG thinks of dogs as being bound together by a hidden essence so that any dog is a dog in virtue of possessing the relevant essence. Put this way, the implication would appear to be that children are, first and foremost, essentialists about categories as opposed to individuals.[4] Essences are conceived of as being hidden and causally responsible for the observable properties of things. Due to this causal connection categorising things

[3] See Ellis, *The Philosophy of Nature: A Guide to the New Essentialism, (*Cheshum: Acumen, 2002) and Mackie, *How Things Might Have Been*, (Oxford: Oxford University Press, 2006) for a more detailed account of this distinction.

[4] Gelman, *The Essential Child*. (New York: Oxford University Press, 2003) is clear on this point.

on the basis of their observable properties will generally result in their being assigned to categories to which they belong. However, such a procedure falls short of being foolproof as, for example, something could appear to be a dog without being a dog and something could appear not to be a dog whilst being a dog. Typically, psychological essentialists regard children as holding a placeholder conception of essence; that is, children do not usually have any substantial views as to the precise nature of the essences of the categories that they adopt an essentialist attitude towards.[5] With respect to the breadth of childhood essentialism there is considerable disagreement. Keil[6] argues that childhood essentialism is restricted to the biological domain. Gelman[7] thinks that children are essentialist about a wider domain of reality that includes the psychological and substances such as water but does not include artefacts. And Bloom[8] holds that children are even essentialist with respect to artefacts such as coffee pots and works of art. What is important to appreciate is that as the psychological essentialist is making a claim about the metaphysical commitments of children she is not thereby committing herself to the truth of essentialism qua metaphysical doctrine.

---

[5] D. Medin and A. Ortony, 'Psychological essentialism'. In S. Vosniadou (ed.) *Similarity and Analogical Reasoning,* (New York: Cambridge University Press, 1989).

[6] F. Keil, *Concepts, Kinds, and Cognitive Development.* (Cambridge, MA: MIT Press, 1989).

[7] Gelman, *The Essential Child,* op. cit..

[8] P. Bloom, *How Children Learn the Meaning of Words* (Cambridge, MA: MIT Press, 2000); P. Bloom, *Descartes' Baby: How the Science of Child Development Explains What Makes us Human* (New York: Basic Books, 2004); and P. Bloom, *How Pleasure Works: The New Science of Why we Like what we Like* (London: Bodley Head, 2010).

As advocates of psychological essentialism portray a commitment to essentialism as a deeply entrenched, universal and innate characteristic of children it is natural to regard them as making a substantial claim about human nature: specifically, that it is part of our distinctive human nature to hold an essentialist outlook on the world (at least when we are children).

At this point a comment about the relationship between psychological essentialism and the kind of essentialism discussed by contemporary metaphysicians is in order. Contemporary essentialism about categories is often characterised as a view about natural kinds, where natural kinds are conceived as objective categories the existence and membership of which is independent of human interest and judgment.[9] This immediately implies that essentialism is not a doctrine that applies to types of artefacts. Moreover, it is often said that Darwin's theory of evolution by natural selection implies that essentialism doesn't apply to biological categories. For example, dogs don't have an essence as any property that dogs currently have need not be present in their descendents.

All this might appear to suggest that psychological essentialism, with its frequent references to the biological and the artefactual, is a misnamed doctrine. However, such a view would be a mistake for two reasons. First, what contemporary philosophers who take essentialism seriously are saying is that essentialism isn't plausible with respect to biological and artefactual categories. But it doesn't follow from this that to think of the biological or the artefactual in essentialist terms is incoherent. Thus it becomes an empirical question as to whether children (or anyone

[9] A. Bird, 'Essences and Natural Kinds'. In R. Le Poidevin, P. Simons, A. McGonigal and R.P. Cameron (eds.), *The Routledge Companion to Metaphysics,* (Abingdon: Routledge, 2009).

else who is philosophically or scientifically unsophisticated) are essentialist about dogs, coffee pots, and the like. Second, the notion of essentialism that the contemporary metaphysician operates with seems to be unduly restrictive in only allowing properties such as intrinsic physical properties and their kin to belong to essences. But why can't having a particular history or bearing a specific relationship to the human mind be part of a category's essence given that histories and mental states are as much a part of the natural world as intrinsic physical properties?

Considerable empirical evidence has been presented in favour of psychological essentialism.[10] To get a flavour of this consider Frank Keil's[11] classic experiment. Keil showed children and adults a picture of a racoon. When asked these subjects answered that the picture was of a racoon. They were then told that the pictured animal underwent a series of changes including changes to its appearance (through fur-dying and plastic surgery), the insertion of a smell sac, and modifications to its behaviour. They were then presented with a picture of an animal resembling a skunk and told that it was of the original animal post-modification. When asked about the identity of the animal at this stage children over the age of seven and adults systematically answered that it was a racoon despite its appearance indicating that for them something's being a racoon is a matter of its origins and/or hidden nature rather than its observable properties.

In this paper my concern is to not to evaluate the evidence for psychological essentialism but, rather, to determine the philosophical significance of the doctrine.

---

[10] See S. Gelman, 'Psychological Essentialism in Children', *Trends in Cognitive Sciences,* **8** (2004) 404-409, and S. Carey, *The Origin of Concepts.* (New York: Oxford University Press, 2009) ch 13, for helpful overviews.

[11] F. Keil, *Concepts, Kinds, and Cognitive Development,* op. cit..

The particular philosophical issue that I will focus on is that concerning the relationship between the contents of an individual's mind and the world external to her skull.[12]

3. Externalism

According to externalism the relationship between the contents of an individual's mind and the world beyond her outer surfaces goes beyond the mere causal. Rather, the very identity of the concepts and thoughts she has will depend on the nature of the external world that she is embedded in. Consequently, it is in principle possible for two individuals to be molecule for molecule replicas (or identical in terms of their intrinsic physical properties) yet have divergent concepts and thoughts due to the fact that they inhabit quite different environments. Externalism, contrasts with internalism. Internalists reject the view that there exists this non-casual relationship between the mind and the external world. For them, the contents of an individual's concepts and thoughts supervene upon their intrinsic physical properties so that molecule for

[12] A number of philosophers and cognitive scientists have utilized a commitment of psychological essentialism in addressing philosophical issues. For example, S. Laurence and E. Margolis, ('Radical Concept Nativism' *Cognition* **86** (2002), 25-55) and S. Carey, (*The Origin of Concepts,* op. cit.) employ psychological essentialism in seeking to undermine Jerry Fodor's argument for radical concept nativism. (J. Fodor, *The Language of Thought.* (Cambridge, Mass: Harvard University Press, 1975); J. Fodor, 'The Present State of the Innateness Debate' in his *Representations* (Cambridge, MA: MIT Press, 1981); and J. Fodor, *Concepts*, op. cit.). And J. Prinz, (*Furnishing the Mind* (Cambridge, MA: MIT Press, 2002)) appeals to psychological essentialism in motivating his proxytype theory of concepts.

molecule duplicates would share their concepts and thoughts no matter how much the environments in which they resided diverged.[13]

Over the last thirty years externalism has become near orthodoxy in the philosophy of mind.[14] This is in no small part due to the influence of Hilary Putnam's paper 'The Meaning of "Meaning". At the heart of Putnam's argument is a thought experiment that is usually described along the following lines. In a distant part of our galaxy there is a planet called Twin Earth that is very much like our own planet. On Twin Earth there is a community of individuals who speak a language very much like English, a community that has a member – call him Oscar2 – who is a physical duplicate of Oscar, a fellow who lives here on Earth. Members of both these linguistic communities apply the word 'water' to the local colourless, odourless liquid that falls as rain, fills their rivers and streams, quenches their thirst, and so on, and intend to apply that word only to stuff that is the same liquid as the local 'water'. One significant difference between Earth and Twin Earth is that the stuff they call 'water' on Twin Earth – the colourless, odourless liquid that fills their rivers and lakes, falls

[13] This way of characterizing the debate between externalists and internalists might seem to be problematic as it assumes a materialist or physicalist view of the mind when Descartes, that paradigmatic advocate of internalism, was a dualist. My reply is that this characterization will work for present purposes as most contemporary externalists reject dualism. See K. Farkas, *The Subject's Point of View*, (Oxford: Oxford University Press, 2008) and T. Williamson, *Knowledge and its Limits* (Oxford: Oxford University Press, 2000) for an attempt to characterise the debate in a manner that doesn't presuppose materialism or physicalism.

[14] However there are critics. For example: T. Crane, 'All The Difference in the World', *Philosophical Quarterly* **41** (1991), 1-25; N. Chomsky, *New Horizons in the Study of Language and Mind*, (Cambridge: Cambridge University Press, 2000); G. Segal , *A Slim Book About Narrow Content,* (Cambridge, MA: MIT Press, 2000); A.S. Wikforss 'Social Externalism and Coneptual Errors', *Philosophical Quarterly* **51** (2001) 217-3; and K. Farkas, *The Subject's Point of View*, op. cit..

as rain, quenches their thirst, and so on – has a physical microstructure that differs from that of the stuff that we call 'water'. For, it is XYZ rather than H_2O. In virtue of this difference the English word 'water' has a different extension than that of the Twin English word 'water'; H_2O, and only H_2O, falls within the extension of the former whereas XYZ, and only XYZ, falls within the extension of the latter. Similarly, English sentences containing the word 'water' have different truth conditions than their Twin English counterparts. For example, the English sentence 'water is wet' is true if and only if H_2O is wet whereas the corresponding Twin English sentence is true if and only if XYZ is wet. Due to this difference of extension and truth conditions, the word 'water' has one meaning on Earth and quite another on Twin Earth. And an upshot of this it that the twins, being fully fledged members of their respective linguistic communities, mean different things by the word 'water' (or understand that word differently) despite their physical similarity. This leads Putnam to conclude that the meaning of a natural kind word on an individual's lips is partly determined by the nature of the external world that she inhabits.

Putnam was primarily concerned with linguistic meaning and with undermining description theories of meaning according to which the reference of a term is determined by its sense or intension (where sense or intension is conceived as a matter of a description associated with the term by the individual).[15] However his argument can easily be extended to generate a parallel conclusion about concepts and thoughts. Here is how such an extension might run. We use language to express our concepts and thoughts. For example, Oscar uses the word 'water' to express one of his concepts and the sentence 'water is wet' to express a belief of his that contains

[15] For a helpful overview see D. Braun, 'Names and Natural Kind Terms'. In E. LePore and B.C. Smith (eds.) *The Oxford Handbook of Philosophy of Language*, (Oxford: Oxford University Press, 2006).

that concept as a constituent. Reflecting the linguistic case, due to the nature of his home environment this concept applies to, and only to, H_2O and the belief containing it is true if and only if H_2O is wet. Similarly, the concept that Oscar2 expresses with the word 'water' applies to and only to XYZ and the belief that contains it is true if and only if XYZ is wet. Due to this difference in extension and truth conditions, Oscar's WATER concept and thoughts differ in content from those of his twin. And as concepts and thoughts are classified partly in terms of their content, the twins diverge in their concepts and thoughts.

This is a tale that has been told many times but there are important features of Putnam's reasoning that, following many recent commentators,[16] I have downplayed. As will become clear, it is important to rectify this situation as I will now do. Putnam assumes that the word 'water' (along with 'gold', 'tiger' and 'lemon') is a natural kind term. What makes it a natural kind term is not merely the fact that most of the samples of liquid that members of the English speaking community characterize as 'water' belong to a common natural kind. All those samples also share certain superficial properties and there is in principle nothing to stop there being a word with a meaning such it that applies to something if and only if that thing has certain superficial properties. What is crucial to a word's being a natural kind term is the state of mind of its users; they must have relevant intentions and make relevant assumptions. This is brought out at several points in the 'Meaning of "meaning"'. For example, imagining himself ostensively defining 'water', Putnam[17] writes:

[16] For example, M. Rowlands, *Externalism* (Cheshum: Acumen, 2003) and R. Wilson, *Boundaries of the Mind: The Individual in the Fragile Sciences* (Cambridge: Cambridge University Press, 2004).

[17] 'The Meaning of "Meaning"', op. cit, 225.

> Suppose I point to a glass of water and say 'this liquid is called water' . . . My 'ostensive definition' of water has the following empirical presupposition: that the body of liquid I am pointing to bears a certain sameness relation (say, *x is the same liquid as y,* or *x is the same$_L$ as y*) to most of the stuff I and other speakers in my linguistic community have on other occasions called 'water'.

When discussing the case of 'gold' Putnam[18] writes:

> when Archimedes asserted that something was gold . . . he was not just saying that it had the superficial characteristics of gold . . . ; he was saying that it had the same general *hidden structure* (the same 'essence', so to speak) as any normal piece of local gold.

In fact, Putnam thinks that related intentions and assumptions are in place with respect to words that are normally contrasted with natural kind terms, for example, those, such as 'pencil', that name types of artefacts:

> When we use the word 'pencil' we intend to refer to whatever has the same *nature* as the normal examples of the local pencils in the actual world.[19]

Returning to the case of 'water' one might ask what it is for two samples of a liquid to bear the same$_L$ relation to one another. Putnam's answer is that it is for them

[18] Ibid., 235.

[19] Ibid., 243.

to have 'the same important physical properties'[20] that is, the same physical microstructure. Consequently, given that the samples of liquid that we routinely call 'water' in our world are invariably collections of H_2O molecules the word 'water' in the English speaking linguistic community has a meaning such that H_2O, and only H_2O, falls in its extension. It is important to note that this doesn't require anyone to know that the crucial property of the liquid they interact with is being H_2O. Although this fact is common knowledge nowadays it wasn't known by anyone prior to the chemical revolution of the 18th century.

In sum then, for Putnam, mental states of members of the linguistic community – in the form of intentions and assumptions – play a key role in making it the case that the nature of the external world enters into the meaning of words such as 'water'.

A second important feature of Putnam's account is his claim that meaning has a social dimension in that what a word means on the lips of an individual is inherited from what it means on the lips of other members of her linguistic community. This is reflected in Putnam's invocation of the division of linguistic labour. An individual might not be able to distinguish between beeches and elms but this does not imply that the words 'beech' and 'elm' (along with her underlying concepts) mean the same thing on her lips. For, she is willing to defer to experts with respect to whether a given tree is an elm or a beech. What connects this with Putnam's point about the role of mental states in determining meaning is that he holds that the social dimension of meaning depends upon individual speakers having appropriate intentions and thoughts in general. For, meaning wouldn't have a social dimension if individual speakers didn't intent to mean by a given word what their fellows mean by that word

[20] Ibid, 232.

or if they didn't recognize the existence of experts and intend to defer to their judgement with respect to the application of words.

4. Externalism and cognitive science

Arguments echoing that of Putnam have been developed by Kripke and Burge.[21] What is perhaps a little surprising is how much they have influenced naturalistically orientated philosophers of mind, that is philosophers of mind who see their enterprise as being closely linked to the empirical study of the mind. For, these standard externalist arguments rely upon intuitions and bizarre thought experiments and make little reference to empirical work in psychology and cognitive science. Moreover, much mainstream work on concepts in cognitive science over the last few decades threatens to deliver a different result by implying that Putnam's twins express the same concept by means of 'water'.

The most prominent theory of concepts within cognitive science developed over the last thirty years is the prototype theory.[22] This began life as a reaction to the so-called classical theory of concepts according to which possessing a concept involves knowing or representing necessary and sufficient conditions for falling under the concept. A prototype is a complex mental representation that, rather than specifying

[21] S. Kripke, *Naming and Necessity, op. cit.*. T Burge, 'Individualism and the Mental', In P.A. French, T.E. Ueling Jr. And H.K. Wettstein (eds), *Midwest Studies in Philosophy **IV**,* (Minneapolis, MN: University of Minnesota Press,1979, 73-121).

[22] E. Rosch, 'Principles of Categorization.' In E. Rosch and B. B. Lloyd (eds.) *Cognition and Categorization* (Hillsdale, NJ: Erlbaum, 1978). L. J. Rips, E.J. Shoben and E.E. Smith, 'Semantic Distance and the Verification of Semantic Relations', *Journal of Verbal Learning and Verbal Behavior* **12** (1973), 1-20. J.A. Hampton, 'Polymorphous Concepts in Semantic Memory, *Journal of Verbal Learning and Verbal Behavior* **18** (1979), 441-461.

necessary and sufficient conditions, specifies the characteristics that any item falling within its extension is likely to have. For example, on this view the concept DOG is a complex representation that specifies properties that dogs generally have, properties that something is likely to have if it is a dog. Examples of such properties might be those of having four legs, having fur, having a tendency to bark, and so on. Thus, the DOG prototype constitutes a description of a prototypical or stereotypical dog and grasping the concept DOG is a matter of having this description encoded in one's head. A prototype also includes a similarity metric so that, for example, determining whether an item x falls within the extension of DOG involves employing a similarity metric in order to determine whether x resembles the prototypical dog to a sufficient extent. A Labrador or a Golden Retriever would be a serious candidate for a prototypical dog but, presumably, a Great Dane or a Pekinese would not be. That an individual would categorise a Pekinese, but not a Siamese cat, as a DOG reflects the fact that employment of the similarity metric generates the result that the former, but not the latter, is sufficiently similar to the prototypical dog to fall within the extension of the concept DOG.

Generally speaking, advocates of the prototype theory regard prototypes as being learned on the basis of experience and as referring to properties that are readily perceivable rather than abstract.[23] A closely related view is the examplar theory of concepts.[24] According to this, at the heart of an individual's concept C is a representation of a particular instance (or number of instances) of C encountered by

[23] This point is emphasized by both J. Prinz, *Furnishing the Mind*, op. cit. and S. Gelman, *The Essential Child*, op. cit..

[24] D.L. Medin and M.M. Shaffer, 'Context Theory of Classification Learning', *Psychological Review* **85** (1978), 207-238.

the individual. Accordingly, deciding whether something falls under the concept in question involves comparing it with the exemplars. For example, central to my concept DOG might be a representation of the dog I had as a child so when I seek to determine whether something that I have encountered is a DOG I do so by working out whether it is sufficiently similar to the dog I had as a child.

With respect to externalism, a key point about prototype and exemplar theories of concepts is that they don't sit too happily with that doctrine. For, it would appear that the prototypes or exemplars in the head of Oscar and Oscar2 will be indistinguishable implying that the twins express the same concepts by means of the word 'water' (for, recall, prototypes and exemplars tend to represent observable properties). Thus, if one wants one's view of concepts and thoughts to be empirically motivated then it seems that one shouldn't be too impressed by externalist thought experiments.

One obvious reply to this is to say that psychologists who work on concepts are primarily interested in the mechanisms by means of which we categorise things and the internal processes by means of which we manipulate the representations associated with our concepts. On this front there is no difference between Oscar and Oscar2. Nevertheless, if concepts are involved in delivering us knowledge about the external world then the identity of the external items that they refer to will be of crucial importance and with respect to this Putnam did establish something important. For, he established that no matter how similar Oscar and Oscar2's prototypes are, as they were constructed in response to samples of different types of stuff they support the possession of concepts that diverge in their reference or extension.[25]

[25] Something like this line of thought is presented by Susan Carey, *The Origin of Concepts*, op. cit. who, following Ned Block ('Advertisement for a Semantics for Psychology'. In P.A. French (ed.)

I'm not convinced that prototype and exemplar theories of concepts can be squared with externalism quite so easily. The danger is that the externalist is begging the question when she asserts that Oscar and Oscar2's prototypes where constructed in response to different types of stuff. Of course, Oscar interacted with H_2O and Oscar2 with XYZ. But H_2O and XYZ agree with respect to the properties that figure in the prototypes in their respective heads so one might equally say that those prototypes were constructed in response to the same type of stuff as belonging to the relevant type is a matter of observable rather than hidden properties. In other words, the externalist has no right to regard the twins' concepts as being natural kind concepts. After all, if the prototype theory is correct then determining whether something falls under a given concept will typically done on the basis of a consideration of its observable properties.

It might be objected that the above point ignores that very aspect of Putnam's line of thought that I have sought to emphasise. This is the idea that when ostensively defining 'water' an individual points at a sample of water and resolves to apply the term 'water' only to stuff bearing the $same_L$ to the ostended sample. My response to this objection is that from the point of view of someone who advocates the prototype theory of concepts this represents a mistaken view of how concepts are acquired. Either, a child acquiring concepts doesn't think of what she interacts with in the manner of the individual in Putnam's scenario or if she does her doing so doesn't enter into the nature and identity of those concepts.

Midwest Studies in Philosophy (Minneapolis, MN: University of Minnesota Press, 1986)) endorses a two-factor theory of concepts.

5. Psychological Essentialism Again

It is at this point that psychological essentialism becomes relevant for it offers an empirically motivated theory that challenges key aspects of prototype and exemplar theories of concepts and would appear to sit more happily with externalism. Indeed, as both Gelman[26] and Carey[27] point out, psychological essentialism was partly motivated by Putnam and Kripke's reflections.

Psychological essentialism implies that with respect to many of their concepts children think that the items that those concepts group together share a hidden essence in virtue of which they fall under the concept in question. Thus, for example, falling under the concept DOG is a matter of having the relevant hidden properties rather than having any superficial properties that dogs typically have. Hence, form the child's perspective, something can appear to be a dog without being a dog and something can be a dog without appearing to be being a dog. This is inconsistent with the prototype theory as that theory implies that the concept DOG is such that being a dog is wholly a matter of satisfying a prototype made up of features that are both readily observable and statistically salient in the child's environment.

According to psychological essentialism the relevant hidden properties are often not known by the child who thinks of them as being a matter of how things in the external world really are in and of themselves. This perspective of the child clearly sits happily with that of Putnam as it implies that a child's concepts work just as he supposes concepts like WATER and ELM work. Moreover, it suggests that it is likely that children will defer to experts. For, if a child recognises that she doesn't know what the essence of being a dog is then she will be disposed to defer to someone who

[26] S. Gelman, *The Essential Child*, op. cit..

[27] S. Carey, *The Origin of Concepts*, op. cit..

she takes not to be hampered by such a lack of knowledge. Note also that the psychological essentialist's emphasis on the perspective of the child echoes Putnam's emphasis on the psychological state of the individual ostensively defining 'water'.[28]

6. Psychological Essentialism and Externalism

In the remainder of this paper I will argue that the relationship between psychological essentialism and externalism isn't as clear-cut as I have thus far implied. Rather, psychological essentialism serves to undermine the kind of externalism that is commonplace in contemporary philosophy of mind. This is not to say that psychological essentialism implies that externalism is false; rather, that the way in which the external world determines the contents of our concepts and thoughts is severely constrained and directed by our underlying mental states. Consequently, an individual's mental states play a more substantial role in determining the content of her concepts and thoughts than is recognized by orthodox externalists. In arguing for this conclusion I will tend to focus on natural language words but my reasoning applies just as much to the concepts expressed by such words. I will do this for ease of exposition and to maintain consistency with Putnam's description of his externalism.

To explore the issue I will begin by considering a problem that Devitt and Sterelny[29] raise for a purely causal theory of reference. Recall that one of Putnam's

[28] None of this is to say that the psychological essentialist is compelled to deny the existence of prototypes. For, she can accept that such structures exist and are routinely employed in making categorization decisions on the hoof so long as she resists identifying them with the concepts that they so help deploy.

[29] M. Devitt and K. Sterelny, *Language and Reality,* (Oxford: Blackwell, 1987).

targets was the description theory of reference and (along with Kripke) he is often characterized as wishing to replace such a position with a causal theory. Now, consider an individual pointing at a sample of water and saying 'I'll call that type of stuff "water"' (or, alternatively, pointing at a dog and saying 'I'll call that kind of thing "dog"'). The problem is that the sample or token in question doesn't just belong to the type *water* (or *dog*) but to many others. For example, *thirst quenching liquid, my favorite drink*, *stuff that expands when frozen* (or *pet, mammal*, *vertebrate*). This raises the qua-problem: when the individual points at the sample of water what is it that determines that she succeeds in referring to the sample qua-water as opposed to qua-thirst quenching liquid or qua-my favourite drink, or qua-stuff that expands when frozen. Similarly, what determines that she points at the dog qua-dog rather than qua mammal, qua-pet or qua-vertebrate? Without a convincing answer to this question it would seem that the advocate of causal theory of reference is saddled with the unfortunate conclusion that term like 'water' and 'dog' have indeterminate references. Devitt and Sterelny suggest that the correct response to the qua-problem is to retreat from a purely causal theory of reference and adopt a causal-descriptive theory instead. Accordingly, although 'water' and 'dog' got their reference partly as a result of interactions with samples of water and dogs this fact alone wasn't enough to secure their reference. In addition, the individual ostensively defining these words had an appropriate description in mind: she thought of what she was attempting to name as being a natural kind whose tokens tend to have particular observable properties.

One comment on this line of thought is that it seems to cohere well with Putnam's own. That is, he is not arguing for a pure-causal theory as he portrays the individual ostensively defining 'water' as intending to use that word to refer only to samples of

stuff that bear the 'same$_L$ relationship' to the sample she points at. Moreover, he represents the individual as having a 'stereotype' in mind that she associates with the word in question. So his point is not so much to establish a pure casual theory of reference but to undermine the idea that an individual's internal mental states are the sole determinants of the reference and meaning of the words on her lips.

Nevertheless, the qua-problem does gesture towards something that I think is very important with respect to the viability of externalism and its relationship to psychological essentialism. Focusing on the example 'water' what is it to bear the same$_L$ relationship to the ostened sample of colourless, odourless, thirst quenching liquid? As we have seen, Putnam thinks that it has to do with having the same microstructure and in the case in question that would involve being composed of H_2O molecules. That being composed of H_2O molecules is what is needed to bear the same$_L$ relationship to the ostended sample is not knowable a priori according to Putnam, a line of thought which sits happily with place-holder conceptions of psychological essentialism. Rather, it is a matter for science to discover. But this raises a further question: does the individual need to think that to bear the same$_L$ relationship to the sample of liquid before her a sample of liquid has to have the same microstructure? Putnam is committed to a negative answer to this question. This is because he thinks that the meaning of words like 'water' and 'gold' have remained constant over centuries so that they meant just what they mean now at a point in time when no-one had the scientific sophistication to think of the same$_L$ relation in microphysical terms. What I want to suggest is that however this question is answered there are serious repercussions for externalism in the light of psychological essentialism. Thus, the question poses a dilemma for the externalist neither horn of which she should find attractive.

Suppose that, following Putnam, we answer the question negatively in saying that the individual need not think of the $same_L$ relation in terms of microstructure. Let's accept that what makes water water is a matter of microstructure. In other words, that water is essentially H_2O. However, a parallel point could not be made of all types of liquid for which we have concepts. Consider, for example, milk. Any sample of milk will have a particular microstructure and a physical makeup in general. It will largely consist of molecules of H_2O along with various vitimins, minerals and fat molecules. Such physical properties will provide the causal basis for the observable properties of the sample, such as its colour, its taste and smell, how it responds to being heated along with its powers to nourish. But having such physical properties is not what makes the sample milk. In other words, milk doesn't have the same sort of essence as water, that is a microphysical or physico-chemical essence. To see this consider the following thought experiment. On an arid planet a team of super-intelligent robots who have never previously encountered water, synthesise a collection of H_2O molecules that they store in a beaker in their laboratory. These molecules form a colourless liquid that any visiting human would be unable to distinguish from water. Would this stuff be water? I contend that it would even though it has different origins from the water here on Earth and even though it doesn't play anything like the same role in the life of its home planet that water does here. For example, it doesn't fall as rain, fill any lakes or rivers or help sustain the life of any living creature. This is a simple consequence of water's having a microsphysical essence.

Now suppose that the robots take the water they have manufactured and mix it with a range of vitamins, minerals and fats that they have also synthesized so as to make something that is identical at the physico-chemical level to the glass of milk that I have just poured from a plastic bottle in my fridge. They don't drink this liquid and

if they did it would certainly not provide them with any nourishment. Neither did they make it with the intention to provide nourishment for any other things. In fact, they are not in contact with any living things that would be nourished by the liquid. Qeustion: is the liquid they have made milk? My answer is that it is not as what makes milk milk is not its physico-chemical properties per se. Rather, the essence of milk has to do with its origins and function; that it is manufactured in the body of a living creature with the function of sustaining and nourishing its young offspring. In short, the milk-like liquid the robots manufacture doesn't have the relevant origins and function to be milk.

Now consider Twin Earth where the liquid that they call milk – a liquid that is produced in the bodies of the creatures they call 'mammals' and is made and used to provide nourishment for the young offspring of those creatures – is largely made up of XYZ. Question: is this liquid milk? I would deliver an affirmative answer on the basis that it has a relevant origin and function.

In sum then, a sample of liquid can fail to be milk whilst being identical at the physico-chemical level to the milk in my glass and something can be milk whilst being very different at the physico-chemical level to that milk. What this implies is not that milk doesn't have an essence but that its essence isn't microsphysical or chemico-physical; rather it is functional or bio-functional. Neither does it imply that 'milk' isn't a kind term or MILK a kind concept, just that the relevant kind is functional or bio-functional rather than physical. Some philosophers might baulk at this suggestion that 'milk'/MILK is a kind term/concept on the grounds that it groups together items whose behvaour is governed by different physical and chemical laws and distinguishes between items whose behaviour is governed by the same physical and chemical laws. I would respond that they are operating with an unduly restrictive

notion of 'kind' but I don't have to insist on this point for the purposes of my argument. As will become clear, all I need is for my claim about what makes milk milk to be true.

Both the terms and concepts 'water'/WATER and 'milk'/MILK are prominent in our linguistic and mental lives and it is important that a child acquires them early in her development, something that a typical child can be expected to do. Now imagine an individual pointing at a sample of milk whilst saying 'milk' alongside the intention to use that word in future only to refer to stuff that bears the $same_L$ relation to the stuff before her. What meaning will she have bestowed upon 'milk'? What concept will she have acquired? Will it be *milk*/MILK or some orthogonal physico-chemical concept? Echoing the kind of scenario highlighted by Devitt and Sterelny[30] the sample of liquid before her falls both under the concept MILK and under some distinct chemico-physical concept. Earlier I posed a dilemma and we are now investigating the first horn of that dilemma. This involves following Putnam in committing oneself to the view that the individual doesn't think of the $same_L$ relation as being a matter of sharing a common-microstructure with the ostended sample. Rather, she has a neutral or unarticulated idea of the relation. But this gives rise to an indeterminacy problem: why would the naming ceremony privilege the attribution of the meaning *milk* to 'milk' (and the acquisition of the concept MILK) rather than an alternative meaning relating to physico-chemical properties? Here, unlike the kind of cases that Devitt and Sterelny discuss, appeal to a stereotype or in-head description relating to observable properties won't help to disambiguate the pointing act. Let me explain why. When the individual points at a dog she is also pointing at a mammal. So what meaning is attributed to the word 'dog' at the naming ceremony? Is it *dog* or

[30] M. Devitt and K. Sterelny, *Language and Reality,* op. cit..

is it *mammal*? It seems that the stereotype or description in the head of the individual settles this question at least to the extent that it rules out *mammal*. For the description will refer to properties that dogs tend to have but that mammals in general don't have so that the description or stereotype will 'fit' dogs in general in a manner in which it won't fit mammals in general. Another way of putting this is to say that the description would serve in the identification of detection of dogs but not of mammals in general. As we have seen, to say this fits well with Putnam's picture. But in the case of the sample of milk such a move won't help. Any stereotype or in-head description will fit the physico-chemical kind just as much as it fits milk as anything that is like the ostended sample at that level will share the kind of observable properties that will figure in the stereotype or in-head description.

The upshot of this is that if the individual operates with an unarticulated notion of the nature of the $same_L$ relation then she is not going to be successful in attributing a determinate meaning or reference to 'milk' or in acquiring a determinate concept when she attempts to bestow meaning on that word.

But the same holds with respect to 'water'. The sample of water ostended in the will fall under a concept that binds together samples of liquid that have a common origin, 'lifestyle' and role in human life and life in general. One might describe this as the concept of a liquid that fills rivers and streams, falls as rain, comes out of taps, and is fundamental to the survival of most living things. Earlier I argued that MILK is a bio-functional concept. With respect to the concept I am now describing, it might be described as a functional concept. Call this concept FWATER. In the environment of the individual ostensively defining 'water' everything that falls under FWATER also falls under the concept WATER and vice versa. However, the concepts are not co-extensive as the XYZ on Twin Earth falls under FWATER though it is not water.

And the H_2O synthesized by the super-intelligent robots described above falls under WATER but not FWATER.

So the question is this: why does 'water' get attributed the meaning *water* rather than *fwater*? I don't see how any plausible answer can be given to this question if one holds onto the idea that the individual in the naming ceremony operates with an unarticulated notion of the same-L relation.

So far I have focused on the use of ostensive definition to bestow meaning on a word. But ostension can also be used to teach the meaning of word that already has its meaning fixed to another individual. Consider an individual who means *water* by 'water' attempting to teach the meaning of 'water' to someone else by means of an ostensive definition. If what I have said so far is true then for this attempt at teaching to be successful the would-be learner must have an appropriately articulated understanding of the $same_L$ relation in her mind. Without this there will be no fact of the matter as to whether she comes to attribute 'water' the meaning *water* or *fwater*.

Is the second horn of the dilemma any more promising? Taking this horn involves attributing the individual ostensively defining a word such as 'water' (or attempting to learn the meaning of such word on the basis of an ostensive definition) a more fully articulated notion of the $same_L$ relation, where she thinks that bearing that relation to the ostended liquid involves having the same microstructure. There are a couple of worries with this suggestion. The first relates to the plausibility of the idea that when the word 'water' first entered the language it did so as the result of a naming ceremony involving an individual who thought of the liquid she was pointing at as having a microphysical essence. As such an event would have had to have taken place considerably before the scientific revolution of the eighteenth century one might reasonably doubt that anyone operated with such a thought.

The second worry is this. Perhaps the individual will succeed in bestowing the meaning *water* on 'water' and on acquiring the concept WATER but she runs the risk of bestowing a meaning other than *milk* on 'milk' and so not acquiring the concept MILK. Given that milk and water have quite different essences the individual will need to desist in thinking of the $same_L$ relation as being a matter of microstructure in the case of 'milk'. Instead, she would need to operate with an alternative (yet still articulated) idea of what the $same_L$ relation amounted to in the case of milk, one that characterised that relation in terms of bio-functional role. One might wonder why the individual would be motivated to regard the sample of milk so differently from the sample of liquid in operating with different notions of the $same_L$ relation with regard to them. After all it is not as if the milk doesn't have a microstructure or that the water doesn't have distinctive origins and a particular important role in our lives. Nevertheless, the second horn of the dilemma does seem to be preferable to the first for it does explain how 'water' and 'milk' could have come to mean what they mean and how an individual could learn the meanings of these words on the basis of being given an ostensive definition .

What are the implications of this for the viability of externalism? The mere fact that the individual ostensively defining 'water' is interacting with a sample of water does not ensure that she will bestow the meaning *water* on the word 'water' or succeed in teaching that meaning to anyone else. She could just as well bestow or teach the meaning *fwater* (along with her doppelganger on Twin Earth). For the microstructure of the ostended sample of water to have a semantic significance it must be thought of in a relevant way by both the definer and the learner. They must think of the ostended sample as having a physical microstructure and intend to apply the word 'water' only to samples of liquid that have that very microstructure. In other

words, they must have an articulated notion of what the $same_L$ relation amounts to in this case. But they must also have at their disposal alternative notions of the $same_L$ relation that they utilize when dealing with words such as 'milk'.

None of this implies that externalism is false for one can construct a twin scenario where an individual on Earth attempts to bestow meaning on the word 'water' operating with the relevant articulated notion of the nature of the $same_L$ relation. Here the word water will acquire the meaning *water* and she will acquire the concept WATER. Her twin on Twin Earth, operating with just the same notion of the $same_L$ relation, will bestow a different meaning on 'water' and acquire a different concept. However, the resultant externalism will be somewhat chastened as the implication of my reasoning is the that extent to which the external world shapes the meaning of one's words and the content of one's concepts is very much constrained and directed by one's internal mental states. For inhabitants of Earth and Twin Earth respectively to mean different things by 'water' (or express different concepts by means of that word) they must have quite specific mental states lying behind their interactions with the external world, mental states that the external world itself doesn't guarantee that they have. This serves to undermine the kind of externalism that dominates contemporary philosophy of mind and language according to which the mere fact that the samples of liquid that we interact with and label 'water' implies that that word refers to, and only to, H_2O and that H_2O, and only H_2O, fall under the concept expressed by that word. It also serves to undermine Putnam's position even though he emphases the importance of mental states in contributing to the determination of meaning.

So far I have focused on the case of language but my reasoning applies just as much to concepts. Thus, for an individual to acquire the concept WATER from her

interactions with water (be those interactions direct or mediated by her experiences of her fellows' use of the word 'water') she needs to think of the target concept as grouping together items in virtue of their having a common microstructure. If she thinks of the target concept in some alternative but equally articulated way then she will acquire not WATER but some other concept (FWATER, perhaps). And if she is neutral on the question of what binds together the items that fall under the target concept then she runs the risk of failing to acquire a determinate concept.

I now want to consider a potential objection to my line of argument. This draws upon essentialism as a doctrine about particulars as opposed to categories. The idea is that although the sample of liquid that figures in the ostensive definition of 'water' is both water and fwater it is essentially the former and only contingently the later. It is this difference that explains why 'water' has the meaning *water* rather than *fwater* bestowed upon it. Thus there is no need to demand of the individual that she has a richly articulated notion of the same_L relation.

I have three points to make in response to this objection. First, it runs the risk of making it a mystery as to how 'milk' means what it does and how we acquire the concept MILK. For, if the essence of a sample of water relates to its microstructure then why doesn't the essence of a sample of milk? One might respond by saying that the essential function of milk relates to biology (that it is produced within the bodies of biological systems for the use of their offspring) so making milk a biological kind and so something in the scientific domain. Whereas, fwater isn't a biological kind but more of an artefactual kind so falling outside of the scientific domain. However, I'm not convinced by this as the essential function of fwater partly relates to its usage by biological systems whose survival depends upon it and which have evolved to utilize it. So why isn't fwater a biological kind? Moreover, if it is conceded that FWATER is

an artefactual concept it might be pointed out that some prominent psychological essentialists[31] argue that our innate essentialist commitments cover the artefactual so that there is an empirical basis for thinking that a particular can have an artefactual essence.

My second point is that such an essentialism about particulars is hardly mandatory. Thus the advocate of this objection needs to produce some justification for it. Such a justification isn't going to come from developmental psychology as psychological essentialists are quite clear that our innate essentialist commitments relate to categories and kinds and not particulars. Thus, there is little empirical support for the claim that it is part of our innate metaphysical perspective on the world that particulars have essences. And even if it were that wouldn't be much help given that, as indicated in the previous paragraph, psychological essentialists often argue that our essentialist commitments spread beyond the domain of physics and biology. This implies that if empirical work in developmental psychology is invoked to settle the issue there is the real prospect that it will support the claim that from the perspective of the typical human the essence of a particular thing that is an artefact relates to its being an artefact as much as its falling under any kind recognized by science.

A third point is that we have to make sense of how all of the words that we use mean what they mean, of why all of our concepts have the content that they have. As we have plenty of words that refer to artefacts and as the acquisition of many artefactual concepts is fundamental to a child's development the advocate of the objection under discussion runs the risk of making a mystery of how we could have such words and concepts.

[31] For example, Paul Bloom, *Descartes' Baby*, op. cit..

I can envisage another objection to my line of thought that runs as follows. Perhaps it is correct to say that a range of articulated notions of essence are needed to ground the meanings of the words of our language and the concepts that we use them to express. But it doesn't follow from this that every individual need have and employ such a range of articulated notions of essence. For we mustn't forget that one of Putnam's key points relates to the social dimension of meaning. It is only the individual members of the linguistic community who first coin a word – or the experts with respect to the application of that word – who need to have and employ the relevant articulated notion of essence (be it microstructuaral, bio-functional, or whatever).

I have two replies to this objection. The first is that it makes it too easy to know the meaning of a word or grasp a concept and rules out as impossible perfectly normal phenomena such as failing to understand a word and misunderstanding a word. Being a competent member of a linguistic community doesn't imply that one knows the meanings of all the words of the community's language or grasps all the concepts expressed by those words. Suppose an individual has encountered the word 'vitamin' but knows little about what vitamins are. Then they could can hardly be said to grasp the concept VITAMIN or mean *vitamin* by that word. In such a case the individual could be expected to be aware of their ignorance so as not to make any claims about understanding the word or concept in question. But there are other cases where an individual mistakenly believes that she knows what a particular word means in the wider community. A common example relates to the word 'disinterested'. Many people think this word means *uninterested* rather than *unbiased.* If such a person were to describe someone as 'disinterested' they would be saying that they were uninterested rather that unbiased. This would be so regardless of the meaning of the

word in the wider community and even if the individual intended to mean by 'disinterested' just what everyone else meant. In sum then, the familiar phenomena of failing to understand the meaning of a word and misunderstanding a word that one uses suggest that for an individual's linguistic knowledge and concepts to line up with those of her fellows considerable demands are placed on her underlying mental states.

None of this is to deny the existence of a division of linguistic labour. Suppose I can't tell elms from beeches. I can till mean different things by the words 'elm' and 'beech' and mean what the experts mean by them. But that this is the case requires me to meet various conditions. I know that 'elm' and 'beech' name distinct species of trees and so employ the concepts TREE and SPECIES in connection with those words. I think that the respective species picked out by 'elm' and 'beech' are different in ways broadly similar to those in which oaks and sycamores (which I can tell apart) differ. Hence, I think they differ with respect to leaf shape, size, DNA, evolutionary history, and such like. I also accept that there are experts and would defer to them but I have some idea about what makes an expert an expert, how to find one, and the kinds of techniques they would use. So it would seem that the divide between me and the experts isn't so extreme and that to avail myself of the division of linguistic labour I have to have quite a rich body of specific mental states.

My second reply is that the objection doesn't sit very well with practice in developmental psychology where it is taken as a real possibility that individuals undergo conceptual development as they mature. Thus, for example, a developmental psychologist might argue that the concept that a typical five year old child expresses by means of the word 'cause' differs from that expressed by a typical ten year old when she uses 'cause', which in turn differs from that expressed by a typical adult when she uses 'cause'. But if the power of the wider linguistic community to enter

into the mind of the individual is as great as the objection implies, then such conceptual development is an impossibility. But rejecting the coherence of orthodox developmental psychology seems to me to be too high a cost of endorsing the objection under consideration.

7. Conclusion

In this article I have argued that, despite first appearances, psychological essentialism undermines the kind of externalist view of the content of our concepts and thoughts that has become commonplace in the philosophy of mind and language. This a consequence of the psychological essentialist's emphasis on a range of concepts that includes those of biological phenomena and artefacts as well as those of types of physical stuff. If a child is to acquire such a wide range of concepts then she will need to have at her disposal a range of articulated notions of essence and bring the relevant notion of essence to bear in each particular case. For example, the articulated notion of essence that she will need to deploy in acquiring the concept WATER will be different from that that she needs to deploy in acquiring the concept MILK. Without the appropriate articulated notions of essence a child will not acquire these concepts no matter how much water and milk she interacts with. This is not to say that the extra-cranial world plays no role in determining the contents of our thoughts and concepts but the extent to which it does is severely constrained and directed by our internal mental states.