PUTNAM'S TRADTIONAL NEO-ESSENTIALISM

NEIL E. WILLIAMS (University at Buffalo)
forthcoming: ***Philosophical Quarterly***

Abstract — *Recently, a number of views have appeared in the literature defending what might be called 'neo-essentialism' about natural kinds. These views are otherwise essentialist, but purport to break from (and improve upon) the traditional essentialism of Kripke and Putnam by rejecting the claim that essences must be comprised of intrinsic properties. However, I argue that this so-called break from traditional essentialism is not a break at all, as the widespread interpretation of Putnam according to which he takes essences to be intrinsic is mistaken. Though it can be argued that Kripke's views are consistent with the standard interpretation, the same cannot be said of Putnam's. Putnam makes no claim to the effect that essences must be intrinsic, and offers at least one example of an essence that is non-intrinsic. I conclude that Putnam's traditional essentialism has been misinterpreted, and consequently that neo-essentialism is not so neo after all.*

1. ESSENTIALISM OLD AND NEW

"*Essentialism*," we are told, "is the view that natural kinds are individuated by essences, where the essence of a given natural kind is a set of intrinsic (perhaps observable) properties, each necessary and sufficient for an entity's being a member of that kind."[1] Definitions of 'essentialism' like this are commonplace in the literature on natural kinds and species. Whether the depiction is detailed or brief, *traditional* natural kind essentialism—the view popularized by Kripke and Putnam in the 1970's—tends to be characterised as the view according to which:

> [1] Members of a natural kind possess an essence that is both necessary and sufficient for membership in the kind.
>
> [2] (In conjunction with the laws of nature) the essence contributes to the production of the superficial properties of a sample of the kind, some of which are typically associated with the members of the kind.
>
> [3] The essence of a natural kind is a set of intrinsic properties.[2]

---

[1] Wilson (1999: 188).

[2] This basic characterization, including explicit statements of [3], can be found at: Boyd (1999: 146), Ellis (1999: 68, 2001: 70), Ereshefsky (2007), Ereshefsky and Matthen (2005: 16), Lange (2007: 373), Machery (2005: 447), Okasha (2002), Slater (2005), Sober (2000 :149), and Wilkerson (1986: 64, 1995: 32, 1998: 225). This list is the product of a brief search, and is not intended to be exhaustive. I am confident that further digging would turn up additional references. In fact, this characterization of essentialism is so common in the literature, and presumed in so many of the connected discussions, that I have no hesitation describing it as a matter of *philosophical lore*. (A brief informal survey—that is, I asked a few colleagues—confirms this claim.)

As a typical example, a natural kind such as gold counts as a natural kind because all and only members of the kind possess the particular essence of gold; that is, being composed of atoms all of which have the atomic number 79. In virtue of possessing this particular essence (in conjunction with the laws of nature), we can explain why samples of gold tend to be shiny, malleable, dense, and so on. This essence, being composed of atoms all of which have the atomic number 79, is an intrinsic property.

Recently, a number of views have appeared in the literature defending or describing what might be called '*neo*-essentialism' about natural kinds. These views are essentialist to the extent that they endorse [1] and [2] (or something very much like them), but purportedly represent a break with traditional essentialism because they reject [3]. For instance, Griffiths endorses a form of neo-essentialism he calls "relational essentialism" which replaces [3] with a more permissive criterion that allows intrinsic *or* relational properties to be essences.[3] And Okasha argues that essentialism is much easier to maintain in the face of biological evidence if [3] is dropped, claiming that Kripke's and Putnam's "error" comes from "accepting an account of the purposes of classifying which implies that relationally defined kinds cannot have fundamental scientific importance," and further that many anti-essentialist arguments against species can be met if essentialism is not "defined by essential *intrinsic* properties."[4] In fact, Okasha claims that the biological counterexamples to [3] constitute "what philosophers of biology call the failure of essentialism."[5] In a similar fashion, Sober regards the biological counterexamples as responsible for the "death of essentialism."[6] In each case the suggestion is the same: essentialism can be greatly improved if [3] is replaced with a more liberal criterion that permits relational properties alongside the intrinsic ones.

However, despite all this recent kerfuffle, I do not think that the third criterion is part of the traditional essentialist treatment of natural kinds as defended by Putnam. Though Kripke's views seem to fall in line with the standard interpretation (*prima facie* at least), Putnam's do not. The essentialism defended by Putnam's bears no commitment to [3], and the examples he employs indicate that he is open to non-intrinsic essences. Therefore to the extent that there is a single view that we can call 'Kripke *and* Putnam's traditional essentialism' that best incorporates the views of each (and not merely their overlap), it is not captured by the interpretation above.

To be clear, my interest is with the correct interpretation of Putnam's views, and not Kripke's. The only reason that talk of Kripke arises at all is that both Kripke and Putnam are credited with defending the 'traditional' view, and

---

A word of caution: the above authors characterize essentialism in the way described, and typically utter 'Kripke' or 'Putnam' in the next breath (or borrow their examples), making it clear that they take Kripke and Putnam to hold this sort of view, but it is possible that they take 'traditional natural kind essentialism' to name one thing, and 'the view held by Kripke and Putnam on natural kinds' to name another, however similar. That said, I think philosophical lore is that: traditional natural kind essentialism is characterized as above; Kripke and Putnam are traditional essentialists; and Kripke and Putnam hold the view characterized above. (And probably even that traditional essentialism is characterized the way it is *because* it is the view held by Kripke and Putnam.)

[3] Griffiths (1999: 218). Based on his examples, the 'or' appears to be exclusive, but Griffiths does not elaborate. See also LaPorte (2004).

[4] Okasha (2002: 210).

[5] Okasha (2002: 205).

[6] Sober (2000: 148).

characterisations of it are typically presented as the view that each holds. Rarely is any effort made—if any is at all—to show how their views differ, if it is even appreciated that they do.[7] Regardless, whatever the details of Kripke's essentialism happen to be, I argue that the standard characterisation of 'traditional' essentialism does not capture Putnam's natural kind essentialism. Consequently the characterisation of traditional essentialism is likewise mistaken. Additionally, this claim is intended to cover just the interpretation of 'traditional' essentialism as thought to be held by Kripke and Putnam, and not any 'ancient' or 'classical' versions of essentialism as held by Aristotle, Locke, Leibniz, Mill, and so on. It is certainly possible that these historically prior essentialists endorsed (perhaps explicitly) something like [3], and if this is the case it suggests one reason why Kripke and Putnam are interpreted as holding the same, but the interpretation is mistaken nonetheless.

If I am correct, this means that the so-called 'failure' of essentialism does not apply to traditional essentialism, and that neo-essentialism is not so 'neo' after all. It is good old traditional essentialism of the sort defended by Putnam. To put the point differently, the standard characterisation of traditional essentialism is mistaken; [3] should be replaced by something like the following:

> [3*] The essence of a natural kind is a set of (relational or intrinsic) properties.

Let me unpack this just a little.

Without settling on anything like a definition for 'intrinsic' (it is a notoriously recalcitrant concept), we can follow Lewis in thinking that intrinsic properties are either those shared by perfect duplicates, or those that something has in virtue of the way it itself is, and not due to some relation (or lack thereof) to other things outside it.[8] It follows that certain relational properties, such as that of having a greater number of teeth than fingers, can be both relational *and* intrinsic, because they involve a relation but would be shared by perfect duplicates. However, though it is possible for a property to be both relational and intrinsic, the assumption in [3*] is that the relevant relational properties are such that they are non-intrinsic, otherwise it would be an uninteresting amendment. (As on this reading a property's being relational is for it to be non-intrinsic, the bracketed clause in [3*] is redundant; nevertheless I have included it to make explicit the manner in which [3*] differs from [3].) The newly included properties are those that obtain just in case the sample stands in some specific relation to something beyond it. For instance, in the case of certain biological kinds, it may be that kind membership is a matter of standing in a specific historical relation, perhaps to an ancestor. An animal that failed to stand in this historical relation would not be a member of the kind, even if it was a perfect duplicate of an animal that was. (If needed, imagine that the non-member duplicate was spontaneously generated.) Such relational essences are thus non-intrinsic.

My argument that Putnam's essentialism is not committed to [3], and is instead committed to [3*], is based on a few key examples of natural kinds that Putnam presents, but that have been largely ignored in the natural kinds

---

[7] Hacking (2007) is a rare and welcomed exception.

[8] See Lewis (1983 and 1986b).

literature.[9] These examples concern the essentialist treatment of *disease kinds* and their proposed essences. To aid in the exposition of the argument, I will first present it very quickly, then again in more detail.

In brief, Putnam endorses the possibility that natural kinds of disease could turn out to have (or do in fact have) a *cause* as their essence. In the case of multiple sclerosis (MS), though no cause has yet been discovered, Putnam suggests that it could turn out an etiologic agent such as a virus could be the cause—and this would be the essence of the disease kind. Though not all disease types turn out to have a hidden structure (some have paradigm cases that are merely constellations of symptoms), others—including tuberculosis (TB) and polio—have a common hidden structure in the form of an etiology, and this is their essence. In the case of TB it is infection by the *Mycobacterium tuberculosis* bacterium; for polio it is infection by the poliovirus. These causes are distinct from the diseases they produce; they are external to them, and are neither identical with, nor contained within, the diseases they cause. Consequently each disease instance stands in a relation (the causal relation) to its cause. As these causes are partly, if not entirely, extrinsic to the disease instances, the causal relation can be translated into a (non-intrinsic) relational property of the disease instances, from which it follows that the essence of certain natural kinds are relational (and non-intrinsic). That was the brief version. Now let me go through it again, only a little more slowly this time.

However, before turning to the argument, a brief warning is in order. My aim is to show that the standard interpretation of Putnam's natural kind essentialism is mistaken. My sole concern is deciphering the view he holds, not what arguments he may or may not have in support of his view, nor whether any such arguments are at all successful. It has been argued by Salmon (and others) that no sound inference can be made from the theories of how extensions for natural kind terms are fixed to the essences of natural kinds, something which Putnam appears to have attempted.[10] Nonetheless, any failed attempt still provides us with a good source of information about Putnam's metaphysical theory of natural kinds. (Unfortunately Putnam does not say much about natural kinds and natural kind essentialism that is not tied up with a theory of reference, so we are forced to read between the lines; if we were not permitted to do that, we would have virtually nothing to go on.) Hence I caution the reader against misinterpreting my use of these arguments as serving any purpose but to help clarify Putnam's natural kind essentialism.

I should also note that inasmuch as the so-called 'failure' of essentialism arises from counterexamples to [3] based on (non-intrinsic) relational properties, modifying [3] in the right way avoids this 'failure'. However, I suspect the 'failure' of essentialism has as much to do with counterexamples to [2] as to [3], so will apply despite my recommended fix of [3]. In what follows I will not be concerned with the 'failure' of essentialism or specific versions of neo-essentialism per se; my aim is just to demonstrate that [3] is a mischaracterization of Putnam's essentialism.

---

[9] All references to Putnam cite pages from his collected works (*Mind, Language and Reality - Philosophical Papers Vol. 2*, 1975; *Realism and Reason - Philosophical Papers Vol. 3*, 1983; and *Realism with a Human Face*, 1990); the individual works discussed are: "Dreaming and 'Depth Grammar'" (1962); "Brains and Behaviour" (1963); "Is Semantics Possible?" (1970); "The Meaning of 'Meaning'" (1975); "Reference and Truth" (1980); and "Is Water Necessarily H_2O?" (1990).

[10] Salmon (1982). See also Salmon (1979), Mellor (1977) and Mackie (2006).

2. ESSENCE AND DISEASE KINDS

In presenting his account of the meanings of natural kind terms, Putnam tells us that membership within a natural kind is not simply a matter of displaying certain superficial characteristics that we tend to associate a priori with the kind, but is instead a matter of having the same *hidden structure*, discovered a posteriori, and that this is the essence of the kind. For instance, for something to be a member of the natural kind we call 'gold' it is not enough—nor is it required—that it be shiny, malleable, precious, and so on (it is neither enough nor required that it displays the superficial characteristics in the 'stereotype'); what matters is that the sample in question bears a certain *sameness relation* to most of the samples that expert members of our linguistic community tend to call 'gold'. (This might mean that many of the samples will reliably display some or all of the stereotyped characteristics after all, but their doing so is not—strictly speaking—criterial for membership.) We are told that this sameness relation is a 'theoretical' relation: it is an empirical question whether or not the sample in question is the same substance as that we call 'gold', the answer to which is to be determined through scientific investigation. Furthermore, Putnam tells us that satisfying the sameness relation is a matter of having the same hidden structure; that is, it is a matter of sharing the same *essence*. He writes:

> In the view I am advocating, when Archimedes asserted that something was gold (*χρυσός*) he was not just saying that it had the superficial characteristics of gold (in exceptional cases, something may belong to a natural kind and *not* have the superficial characteristics of a member of that natural kind, in fact); he was saying that it had the same general *hidden structure* (the same 'essence', so to speak) as any normal piece of local gold (1975: 235).

And also that:

> if there is a hidden structure, then generally it determines what it is to be a member of the natural kind, not only in the actual world, but in all possible worlds. Put another way, it determines what we can and cannot counterfactually suppose about the natural kind (1975: 241).

Central to his discussion of hidden structures and essences are the 'Twin Earth' thought experiments, used by Putnam to draw out various intuitions concerning what we take to be counterfactually true of natural kinds (from which he derives his arguments for meaning externalism). In the scenario most relevant to our purposes, Putnam presents the reader with a far off planet—'Twin Earth'—which is like the Earth in most respects, but where the wet, clear, potable liquid found in most lakes and rivers (to wit, the Twin Earth stuff that answers to the stereotype associated with the term 'water') is not H_2O, but some other compound whose complex chemical formula is conveniently symbolized 'XYZ'. Putnam argues that despite its superficial similarity, we would not take XYZ to be water, because it differs at the chemical level. To be water it would have to be the same stuff as the water 'around here', which we know to be H_2O. From this Putnam concludes that our picture of natural kinds is one according to which displaying the

properties in the stereotype is not sufficient for sameness, instead it is having the underlying nature that matters, and that this constitutes the essence of the kind.

Turning now to the case of disease kinds, the lesson about essencess means that in those cases where there is some hidden structure, the superficial characteristics—the symptoms—are not what determine membership in the kind. According to Putnam, two disease instances could, in principle, display the same symptoms and yet be members of different disease kinds; conversely, two members of the same disease kind could present themselves in entirely different ways. Or at least such is the case for those disease kinds with hidden structures. As Putnam points out, there are a variety of cases where we might be on the same epistemic footing regarding a disease, but that differ as follows: some disease terms name disease kinds that have hidden structures, but others name diseases that have turned out not to and are re-identified via clusters of symptoms. For convenience I shall take it as understood that Putnam is aware that there are diseases of both varieties, and will speak only of those disease kinds that (at least putatively) have hidden structures, ignoring the cluster cases. Hence, for the cases that are of interest to us, to be of the same disease kind two disease instances must share a common hidden structure. So just what are these hidden structures?

In the majority of Putnam's examples (lemons, water, aluminium, elms, gold, tigers), the hidden structure is an *internal* structure; it is some property intrinsic to the members of the kind. In the case of the chemical kinds this hidden structure is a *micro* structure: the hidden structures are scientifically discovered underlying physical characteristics, such as chemical formula (H_2O for water) or atomic number (79 for gold; 13 for aluminium). Putnam writes:

> To be water, for example, is to bear the relation same$_L$ to certain things. But what is the relation same$_L$?
>
> *x* bears the relation same$_L$ to *y* just in case (1) *x* and y are both liquids, and (2) *x* and *y* agree in important physical properties. … Importance is an interest-relative notion. Normally the 'important' properties of a liquid or solid, etc., are the ones that are *structurally* important: the ones that specify what the liquid or solid, etc., is ultimately made out of - elementary particles, or hydrogen and oxygen, or earth, air, fire, water, or whatever - and how they are arranged or combined to produce the superficial characteristics (1975: 238-9).

In addition to stating that the 'important' properties are normally structural, Putnam indicates that in order for *x* and *y* to both be water they must both be liquids. This latter condition is what Putnam elsewhere describes as a matter of *x*'s and *y*'s 'falling under the same semantic marker', where a 'semantic marker' is a less determinate natural kind term than that at issue. Hence something cannot be water without being a liquid, nor a tiger without being an animal, and so on.

The other feature of the structures is their connection to the surface characteristics, as indicated in [2]. As Putnam indicates, the hidden structures are normally responsible for the production of the superficial characteristics. (It might be more precise to say that they are *partly* responsible, as having such-and-such structurally important properties causally produce the superficial properties in conjunction with the laws of nature and a sample's immediate physical

environment.[11]) Putnam does not spell out the details of how precisely this production of the superficial properties is to occur, except to say that the "normal distinguishing characteristics are 'held together' or even explained by deep-lying mechanisms."[12] I take it that talk of 'mechanisms' is a reference to the causal structure of the world: "it is the atomic composition that determines the law-like behaviour of a substance;" and "A and B are the same substance if and only if they obey the same *laws*."[13] It follows that it is because gold has the atomic structure it does that it looks shiny and is malleable in typical environments (that is, at standard temperatures and pressures). If the environment deviates from the typical, so too may the superficial properties. For instance, typical samples of gold have mild impurities, such as the presence of copper, which give rise to gold's characteristic yellow colour; minus the impurities gold has a whitish/silvery colour. In the case of biological kinds like lemons, tigers, and elms, the microstructure is biochemical: it is the specific DNA. It is this DNA then that produces the familiar superficial characteristics of the plant or animal by directing its growth and development. Hence, though no property in the stereotype is necessary or sufficient for kind membership (because changes in laws or context could result in their absence and other kinds might display just those same properties), it will often be the case that some of the stereotype properties are nevertheless among those superficial properties produced by the hidden structures. That said, instances of coincidence between the superficial and the stereotyped tend to be restricted to the physical properties, as underlying structures have no significant role to play in the production of conventional properties that may be included in the stereotype. That gold is valuable and that tigers are dangerous are not properties that flow from the hidden structures.

Though the received account fits the examples considered thus far, it would be a mistake to reconstruct Putnam's essentialism on the basis of these examples alone, as this list excludes important examples Putnam discusses that we have not yet considered. (In diagnosing the mistaken characterisation of Putnam's essentialism I suspect that undue attention to just these examples may be to blame; they certainly take up the lion's share of the discussion in the literature.) What then are the overlooked examples? For starters, there is the case of acids: Putnam suggests that their essence is that of being proton donors. Though most likely intrinsic, this property is not microstructural (it is a micro property, but not structural), it is dispositional. But more importantly, when it comes to disease kinds, Putnam proposes a strikingly different notion of structure; to wit, hidden *causal* structures:

> There are, in fact, almost continuously many cases. Some diseases, for example, have turned out to have no hidden structure (the only thing the paradigm cases have in common is a cluster of symptoms), while others have turned out to have a common hidden structure in the sense of an etiology (e.g. tuberculosis). Sometimes we still don't know; there is a controversy still raging about the case of multiple sclerosis (1975: 241).

---

[11] See Putnam (1983: 71-75).
[12] Putnam (1975: 139). See also (1990: 68-70).
[13] Putnam (1983: 73, 1990: 68).

Hence to claim that natural kinds are united by a hidden structure is not to demand that the structure must be a microstructure, nor, it would seem, that it must be intrinsic. (For now let us leave the matter of intrinsicness to one side, until we have had a chance to better consider the examples concerning disease; we shall return to it shortly.) What matters is that there is a set of properties (however small) that are found throughout the samples of the kind and that bear the right sort of connection to the manifest characteristics of the kind, as per [2].

Regarding that connection, since it is all but criterial for causes that they explain their effects, there is no difficulty in seeing how a causal structure could be connected to some set of effects, and thus how a disease's etiology could explain the symptoms in a particular disease instance, as well as those that are present in paradigmatic cases and typically associated with the kind. After all, like gold and horses, "[t]here are *objective laws* obeyed by multiple sclerosis."[14] Though the model differs from that of kinds with microstructures, it is still the case that the these causal structures contribute to the production of the superficial properties (the symptoms) displayed by instances of the kind. In its simplest form, the cause of the disease causes the disease, and the disease causes the symptoms; we get production from the causal structures via the transitivity of the casual relation. And even if the transitivity of causation is controversial, that the causal structures have *contributed* to the production of the symptoms is not. Let us examine Putnam's disease cases more closely.

Putnam tells us that in the case of tuberculosis the hidden structure is its etiology; that is, the essence of TB is its cause. Therefore, for a disease instance to be a member of the disease kind we call 'tuberculosis', it must have as its cause the same cause as found in the paradigm cases. In other words, it must be an instance of disease caused by a specific bacterial infection. TB—like so many of the diseases whose natures were discovered late in the nineteenth century (anthrax, syphilis, and cholera, to name a few)—is a disease with an identifiable 'etiologic agent' in the form of a pathogen such as a bacterium or virus. Presumably Putnam would take most of the diseases on this list to be kinds whose essences have been discovered.

As it happens, the actual facts concerning TB are not as tidy as Putnam would have us believe. TB has more than one cause (infection by *Mycobacterium tuberculosis* is the most common cause, but TB can also result from infection by *Mycobacterium bovis*, to name just one alternative), so by his own lights Putnam would have to say that 'tuberculosis' names not one disease kind but at least two. In this respect 'tuberculosis' is like 'jade': with the discovery that the green mineral had two distinct microstructures (jadeite and nephrite) we discovered that 'jade' was the name for two natural kinds, not one.[15] Of course, the extent to which TB is like jade in the relevant respects depends on how bacterial kinds are divided. If *Mycobacterium tuberculosis* and *Mycobacterium bovis* (along with all other TB causing bacteria) are appropriately related, then 'TB' may still name just one kind. Nevertheless, as it is of no consequence to the present discussion to what extent Putnam has the correct picture of disease kinds, we need not concern ourselves with any confusions or errors he might have made.[16] What matters, and the lesson to be learned, is that alongside natural kinds whose essences are microstructural,

[14] Putnam (1983: 71).
[15] Putnam (1975: 241).
[16] Footnote suppressed.

we find Putnam producing examples of disease kinds whose essences are their causes. That his understanding of disease might be mistaken is inconsequential.

This is evident in the second disease kind Putnam discusses: multiple sclerosis. Putnam raises the case of multiple sclerosis as part of another externalist thought experiment, designed to show—contra Malcolm and others—that what we mean by 'multiple sclerosis' is not tied to any prior understanding regarding some set of symptoms:

> Consider the following case: there is a disease, multiple sclerosis, which is extremely difficult to diagnose. The symptoms resemble those of other neurological diseases; and not all of the symptoms are usually present. Some neurologists believe that multiple sclerosis is caused by a virus, although they cannot presently specify what virus. Suppose a patient, *X*, has a 'paradigmatic' case of multiple sclerosis. Then Malcolm's view is that, no matter what we find out later, *X has* multiple sclerosis because that is what we presently mean. In particular, if we later identify a virus as *the* cause of multiple sclerosis, and this patient's condition was not caused by that virus, he *still* had multiple sclerosis (1975: 310).

Putnam does not defend the claim that MS has, as a matter of fact, a viral cause, or even that it has any cause at all. What he claims is that we could come to discover a unique viral cause for MS, and that this possibility demonstrates an important feature of how we understand the term 'multiple sclerosis'. In particular, it demonstrates that our understanding of the term is not tied to a cluster of symptoms, for a disease instance can display the symptoms in the cluster and yet fail to be an instance of MS; conversely, we would still consider a disease instance an instance of MS if it was caused by the newly discovered viral cause, even in the absence of the paradigmatic symptoms. Hence our understanding of the term is tied to being the "same as" some initial disease instances in important respects, not a cluster of symptoms. Whether or not MS actually has such a cause, viral or otherwise, has no bearing on the point Putnam makes, nor does whether or not Putnam *believes* that MS has a viral cause. Nor does it matter for the point I want to make: that Putnam raises the possibility of a unique viral cause for MS as part of an argument against an alternative theory of meaning shows that Putnam takes seriously the possibility that the hidden structure that determines whether a disease instance satisfies the *same disease as* relation can be a cause. Which is to say he has no problem with natural kinds having causes as essences.

The third and final disease kind Putnam mentions is polio. Though the discussion is brief, the point is now familiar: the meaning of the term 'polio' is not tied to a set of symptoms, but some underlying condition with a viral origin. "We observe that, when a virus origin was discovered for polio, doctors said that certain cases in which all the symptoms of polio had been present, but in which the virus had been absent, had turned out not to be cases of polio at all."[17] Furthermore, Putnam contends that this viral cause is criterial for a disease instance's being a case of polio, and that this does not represent a change in the meaning of the term 'polio':

[17] Putnam (1975: 329).

> Some philosophers would prefer to say that 'polio' *used to mean* 'the simultaneous presence of such-and-such symptoms'. And they would say that the *decision* to accept the presence or absence of a virus as the criterion for the presence or absence of polio represented a *change of meaning*. But this runs strongly counter to our common sense (1975: 329).

As additional evidence that Putnam treats the cause of the disease as the essence of the disease kind (for at least some diseases) I offer the two following points. The first is that this view of disease similarity (according to which two disease instances count as members of the same disease kind if and only if they have the same cause) is far from new, and was once quite popular, so it is not at all odd that Putnam would subscribe to it (or at least speak as if it could be true). The view Putnam presents is what lies at the core of the 'germ theory' of disease. According to the germ theory, a patient has a disease when she is infected by a pathogen, where the pathogen produces the disease and the symptoms.[18] Understandably, disease types are then taken to vary with the pathogenic micro-organism that produces the disease instances. Though overly simple and unsuitable for a wide range of diseases (genetic diseases present a serious counterexample), the germ theory is highly intuitive, and marked an important step in the development of our medical knowledge. And despite having its heyday at the end of the nineteenth century, similar accounts are not without their adherents today.[19] That Putnam should endorse it—or something like it—is perfectly reasonable. The second point is that I am not alone in my interpretation of Putnam's disease examples. Achinstein, Goosens, and Lange all read Putnam as having the view that disease kinds are determined by their causes (and that these causes are distinct from their effects).[20]

This completes the first step in my argument. The second is that the causes of diseases are distinct from their effects, and are so in such a way as to generate a relation that is extrinsic to the disease instances. In order to argue that Putnam is best interpreted as holding [3*] and not [3], it is not enough that he includes causes as the essences of natural kinds, he must also believe that the causal relation that holds between the cause of a disease and the instance of the disease is not intrinsic to the disease instance. As indicated above, some relations are intrinsic to their bearers; it must be shown that the relations that determine kindhood for disease kinds are not intrinsic in this sense. However, if Putnam believes that causes are appropriately distinct from their effects, then we are well on our way to showing that Putnam allows for non-intrinsic relational essences for natural kinds.

3. CAUSES OF DISEASES AS DISTINCT AND EXTRINSIC

On the surface of it, showing that Putnam believes causes are distinct from their effects is straightforward. That in an instance of causation the cause is distinct from its effect(s) is common knowledge (among philosophers anyway). Though the aphorism likely dates back even further, Hume famously claimed in the *Enquiry* that "every effect is a distinct event from its cause" (Bk IV). Since then virtually

---

[18] Thagard (1999).
[19] Most notable is Whitbeck (1976, 1977).
[20] Achinstein (1968), Goosens (1977), Lange (*forthcoming*).

every philosopher to write on causation has adopted (implicitly or explicitly) some version of the distinctness criterion. David Lewis—in his highly influential contemporary piece on causation—follows suit, claiming that causation is to be analyzed "in terms of counterfactual dependence between distinct events."[21] And Swain claims that "virtually everyone who works on causation" agrees that for *c* to be properly called the cause of *e*, "then *c* and *e* must be *distinct*."[22]

I do not think a great deal of argumentation is required here. What ultimately matters is not whether causes are *in fact* distinct from their effects, but whether it is reasonable for us to assume that Putnam believes this concerning the causes of disease instances. And given that it is both common knowledge and common sense (and he says nothing to make us think otherwise), we have every reason to think that Putnam would take causes in general to be distinct from their effects, and therefore that he would think the same regarding diseases and their causes.

This reading of Putnam would run into problems only if he took the causes of diseases to be parts of the diseases themselves; that is, if Putnam thinks either that a disease *just is* its cause, or that a cause is a proper part of a disease. Hence demonstrating that the etiological essence of disease kinds is an extrinsic relation requires showing that Putnam treats the cause of the disease as distinct from the disease instance, in both the sense of identity and of proper parthood (partial containment—'overlap'—of the cause and the disease would mean that some portion of the cause was not part of the disease; this would be sufficiently distinct to support the cause's being non-intrinsic in the relevant sense). This will be done in two steps. First I will provide textual evidence that Putnam takes a cause of a disease to be distinct from the disease in terms of their being non-identical. Second I will argue that we can extrapolate from Putnam's claim that the cause is the essence of certain disease kinds—making use of a Putnam-style 'Twin Earth' scenario—that he is committed to the possibility (if not actuality) of instances of disease kinds that do not contain (do not have as proper parts) their causal essences. It follows that if the cause of the disease is distinct from the disease in terms of identity and complete containment, then the cause is appropriately 'external' to the disease instance for the causal relation to be extrinsic to the disease.

In the first case, it might be objected that while there is no error in taking infection by the *Mycobacterium tuberculosis* bacterium to be the cause of tuberculosis, it nevertheless fails to be distinct from the disease, for being infected by the *Mycobacterium tuberculosis* bacterium is *identical* with the disease. Though this strikes me as a plausible view of disease, it is not Putnam's. Putnam is not suggesting that a disease *just is* its cause any more than he suggests that a tiger just is its DNA; in each case the essence is that which determines the *same as* relation to the archetypical instances, and which picks out the members of the kind. We must be careful not to confuse the question of what diseases are (what sort of entity), with that of what makes two disease instances members of the same kind. Recall the quotation presented above (Putnam 1975: 238-9), that for *x* and *y* to be of the same kind they need to agree in important physical properties, but must also both fall under the same semantic marker. In order for something to be a tiger it must have the right DNA, but it must also fall under the semantic marker 'animal'). Similarly, for something to be an instance of MS it must be a disease—that is, a certain state or condition of a person—and it must be a state or condition that results from a

[21] Lewis (1986: 191).
[22] Swain (1980: 155).

specific viral infection. Even if a continued viral infection forms part of the resultant state, this state nevertheless differs from its viral origin. Part of the confusion here may come from things that Putnam himself sometimes says about how the important properties are *normally* those that specify what the members are ultimately made out of; the quotation just recalled being one such case. But this thought that essences involve constituents does not extend to all cases: it is not exceptionless. Putnam's essences are those properties that determine the 'same as' relation between members of a kind; many involve fundamental constituents, but not all. A disease kind can thus have a bacterial infection as its essence without being identical with it.

But more to the point, how could it be anyone's view that the cause of a disease is identical with the effects it causes? This stinks of nonsense. We might be able to make sense of causes that overlap with their effects, but as was pointed out above, for something to be a cause it is necessary that it be distinct from its effects. Putnam could not propose a scenario in which "we later identify a virus as *the* cause of multiple sclerosis" and yet maintain that within the scenario the virus just is MS.[23] The only way to coherently maintain that the cause of MS is identical with the disease would require reading this quotation in such a way that 'multiple sclerosis' names something other than the disease, such as a cluster of symptoms, because otherwise the cause would be its own effect. But nothing could be further from Putnam's view: "by the term 'multiple sclerosis' he *means* whatever *disease* causes such-and-such symptoms, and not just the simultaneous presence of the symptoms."[24] To suggest that Putnam sometimes uses 'multiple sclerosis' to name of a group of symptoms is to have seriously misunderstood Putnam's central point about natural kind terms, and would violate the very account of meaning he is using the example to demonstrate. Without belabouring the point, if Putnam intended to claim that diseases like MS were identical with their causes, he had every opportunity to do so—that he did not is telling. Instead he claims something entirely at odds with this identity claim, namely that terms like 'multiple sclerosis' name diseases, not sets of symptoms, and that a specific virus or bacteria is the cause of the named disease. It follows that Putnam does not take disease instances to be identical with their causes.

What then of containment: does Putnam think that the causes are proper parts of the diseases they cause? Here the story gets slightly more tricky. In each of Putnam's examples of disease kinds it looks as if the cause of the disease (or the speculated cause in the case of MS), is consistent with the cause's overlapping with the disease, and might even be—at a stretch—thought of as a proper part of the disease. For instance, the cause of TB is infection by the *Mycobacterium tuberculosis* bacterium, but continued infection is part of the disease process. And assuming MS has a (typical) viral etiology, the disease process involves replication of the viral genome and continued viral infection. Now, it may be that this impression is misleading, and that—as a matter of fact—neither case involves containment, and so do not constitute even putative counterexamples. After all, though continued infection might be part of a disease process or be constitutive of a disease state or condition, the initiation or onset of the process would nevertheless appear to be a distinct event. Nonetheless, however we interpret the cases, I contend that any putative containment is an artefact of the examples and not a feature of Putnam's account, and hence not a problem. To see this we must compare the actual disease

[23] Putnam (1975: 310).

[24] Putnam (1975: 311, 312).

processes with possible disease processes, applying the sort of counterfactual reasoning about kinds that Putnam himself uses.

Imagine a Twin Earth disease—as common there as TB is here—called 'twuberculosis' (TWB), whose onset is just like tuberculosis (it has the same infectious agent), but where the progression of the disease is quite different. In cases of TWB, after the bacterium enters the body it produces a tiny amount of toxin, and is destroyed in the process. The tiny amount of toxin then uses the body's own immune system to produce more of the same, and spreads about the body. The result is a continued process which is just as bad for the twuberculosis sufferer as the typical process is for the tuberculosis sufferer, though the two have few, if any, symptoms in common. TWB, it would appear, is a disease whose cause (infection by the *Mycobacterium tuberculosis* bacterium) is not part of the disease process. It initiates the process, but unlike TB, continued bacterial infection is not part of the disease. The relevant bodily states of someone with TWB do not resemble those of someone with TB. However, given that TWB and TB have a common cause, it follows, according to Putnam's account of disease kinds, that TWB *just is* TB. They differ in terms of many features, but this is permitted by Putnam's account of kinds; what matters for kind membership is having the same essence, not the stereotype. Hence it is perfectly consistent with Putnam's account of disease kinds that the cause is no part of the disease it causes (even though this may occur in instances of TB around here). Moreover, though TWB is presented as a mere fantasy disease, the facts can be as strange as the fiction. The human body evolves and changes with time, and with it the immune system and its responses to infection. In time, infection by the *Mycobacterium tuberculosis* bacteria could plausibly result in a process just like this. And though at such a time we might express legitimate doubts as to whether this ought to count as the same disease, apt pupils of Putnam should have no such reservations.

A second case, somewhat closer to home, concerns deficiency diseases such as scurvy. Here the cause is not associated with the actions of an etiological agent—a 'something' if you will—but rather a 'nothing': the absence of something required for a typical healthy process to be actualized. The negative effects of scurvy (spots, spongy gums, bleeding and loss of teeth) can be halted quite abruptly by filling the absence through normal intake of vitamin C, so there is a sense in which the deficiency is 'present' throughout the disease process, but it is a stretch to claim that the absence is part of the disease. That absences can be causes is suspicious enough, that they should be *proper parts* of anything is an abuse of the term. Hence, on the assumption that Putnam would extend his account to this sort of disease (and we have every reason to think he would, after all, these are diseases with an identifiable etiology and are not mere clusters of symptoms), we have additional cases of disease kinds whose instances do not include their causes. Hence, Putnam treats causes as non-identical with the diseases they cause, and permits causes that are not proper parts of the disease processes they initiate.

Before we can conclude that Putnam treats the causes of diseases as distinct, and therefore relational in the appropriate (non-intrinsic) sense needed for [3*], we must deal with one final objection. Throughout this section I have assumed a particular ontological picture of disease that Putnam seems to be working with, according to which diseases are states or processes of persons. As such, diseases are dependant entities: they cannot exist independently of the persons who have them. Consequently, it might be objected that in virtue of their ontological status, diseases are not fit to serve as natural kinds at all, and are only *ways* that natural kinds (such as humans) can be. The relevant category might be

that of 'diseased persons'.[25] In that case, even though the cause of a disease instance would be distinct from and (at least partly) external to the disease instance, both would be intrinsic to the diseased person, and so intrinsic after all. If Putnam's ontological treatment of disease is like this, then we have a problem.

The short response is that Putnam's ontological account of disease is not like this, and so the cause's being external to the disease instance generates the sort of extrinsic relation we require. The objection admits of a strong bias towards treating only substances as natural kinds, but this is a bias Putnam does not share. Though they are his only example of non-substance kinds, Putnam frequently speaks of diseases as natural kinds, as we have seen. He clearly has no requirement that the class of natural kinds be restricted to substances alone. However, in order to defend the extrinsic-friendly interpretation I have assumed throughout this section, it will be useful to quickly take stock of Putnam's ontological picture of disease, as this will provide a clearer foundation for seeing how the essences come out as relational and non-intrinsic.

Putnam's ontological picture of disease has three main elements: the cause of the disease, the disease itself, and the symptoms of the disease (which are caused by the disease). Most of Putnam's interest in discussing disease concerns the meanings of disease terms: his central claim is that disease terms name the underlying cause of the symptoms usually associated with the disease term, but are not tied to some specific cluster of symptoms. From this we get two of the ontological elements, the disease and its symptoms. Beyond serving as the cause of the symptoms, we get only a little direction from Putnam regarding what he takes disease instances to be. The best indication is when he speaks of: "a 'disease' in the theoretical sense of, say, virus-caused destruction to nerve tissue," from which it would appear that he takes diseases to be states or conditions of the diseased person. (These states might be dispositional states, dispositions whose manifestations include the symptoms of the disease.) He further claims that: "The extension of 'multiple sclerosis' includes whatever illnesses turn out to be *of the same nature* as the majority of the 'paradigm' cases of multiple sclerosis," which in addition to reiterating the above assertion that Putnam has no problem with diseases as natural kinds, is telling in that it speaks of *illnesses*, indicating that this (or something like it) is the relevant semantic marker for disease kinds, not 'ill persons' or the like.[26]

The third element is the cause of the disease, such as the possible viral cause of multiple sclerosis. The virus and the disease it causes in humans are distinct, yet the latter is individuated on the basis of the former. Sticking with the case of MS, this means that (perhaps) a virus is the cause of the disease, which then gives rise to instances of MS through the right sort of contact. The disease itself is the result of that contact: the state or condition of having damaged nerve tissue, which may lead to such symptoms as muscle weakness, degenerate nerve function, bladder or visual problems, and so on.

The picture that emerges is one according to which the symptoms of a disease can vary wildly (though they tend not to) and yet the disease would still count as a member of the same disease kind just as long as *its* cause is the same. Nonetheless, in certain cases the cause of the disease is not entirely separate from the disease itself, and may be a partial or indirect cause of the symptoms of the disease. After all, causation is typically interpreted as a transitive relation, such that

---

[25] Sulmasy (2005) defends a view of this sort.

[26] Putnam (1983: 71).

causes other than the most proximate cause also count as causes. To the extent that the cause of the disease differs from the disease, it becomes a less proximate cause of the symptoms. If some portion of the cause of the disease should overlap with the disease itself, it can be a more proximate cause of the symptoms. This allows us to make sense of those rare occasions when Putnam speaks in a manner that suggests he is treating a virus as a cause of both the disease and its symptoms, without thereby committing him to the identity of the disease and its cause. For instance, in highlighting the extent to which the symptoms of MS could differ without a change in meaning of the term 'multiple sclerosis', he says that we could find "virus caused symptoms."[27] If it is a viral infection that causes the nerve damage of MS, then by transitivity we can say it is that same viral infection that causes muscle weakness, degenerate nerve function, and so on.

We have seen that Putnam's account of natural kinds makes room for natural kinds that have causes as essences; to wit, he provides examples of disease kinds with etiological essences. It has also been shown that Putnam takes these causal essences to be non-identical with the disease instances they cause, and that his account permits their distinctness even if they should partly overlap. Because of the way Putnam understands what it is to be a disease, this distinctness means that diseases bear relations to their causes that are extrinsic to the disease instances. The final step in arguing that Putnam is best interpreted as holding [3*] and not [3], is to show that this extrinsic relation can be translated into the more familiar talk of essential *properties*.

4. PUTNAM'S RELATIONAL ESSENTIALISM

On the majority of accounts, talk of cause and effect tends to be framed in terms of events. As talk of the essences of natural kinds is framed in terms of those properties that samples of natural kinds must instantiate in order to be a member of the kind, we require some sort of translation from the event-based talk of the cause being distinct from the disease to the property-based talk that allows us to treat this distinct cause as a relational property. (Even when talk of cause and effect is not framed in terms of events, the standard alternatives substitute event talk with that of facts or objects, generally not properties, so the need for translation remains.)

The direct—but reasonable—way to get from the event-based talk of the distinct cause to the right sort of property-talk is to define a property along the lines of 'having such-and-such as a cause'. It ought to be clear that '*is caused by such-and-such*' names a relational property, as it picks out a feature instantiated by an entity in virtue of its standing in a particular relation (the causal relation) to some event, fact, or object beyond it (its cause). If the causal relation obtains, then the relational property in question is present. It is a historical relation, one that speaks to origin. As the cause of TB is infection by the *Mycobacterium tuberculosis* bacterium, it follows that instances of TB stand in the causal relation to the cause (the 'infecting' if you will), and so have the relational property of 'having infection by the *Mycobacterium tuberculosis* bacterium as a cause'. Furthermore, as perfect duplicates could differ with regard to what type of cause they are related to, it follows that the relational property is extrinsic.

---

[27] Putnam (1975: 329).

Nevertheless, might it be objected that in order to stand in this relation the disease must be an event, but this does not match with Putnam's ontological category for diseases? Even if the typical causal relation holds between events, this does not restrict us in terms of what sort of entities disease instances are. After all, it is just as natural to speak of a cause as having produced some event as it is to speak of a cause having produced some state, property, object, fact, or what have you, even if these tend to come about as constituents of the effect event. (Consider: the car accident caused the bending of the fender, but it also caused the fender to be bent, the bentness of the fender, and the fact that the fender is bent.) Hence, there is no barrier to the disease instances having the relational property in question.

Conversely, those who might have reservations over having the cause event serve as one relatum of the relation that supports the relational property might prefer to focus on the etiologic agent that is active in the event. *It*—that is, the single *Mycobacterium tuberculosis* bacterium that is responsible for this case of TB—stands in a specific relation (as the key constituent of the cause) to the particular instance of the disease. This suffices to generate the relational property had by the disease instance, that of standing in a certain relation (the causal relation) to a specific bacterium (that responsible), and that is enough.

In sum, Putnam takes causes to be the essences of disease kinds, and takes these causes to be distinct from the diseases they cause both in terms of identity and proper parthood. As the distinctness of cause and effect can be translated into a (non-intrinsic) relational property of the disease entity, it follows that Putnam provides examples of natural kinds whose essences are (non-intrinsic) relational properties. Consequently, Putnam's natural kind essentialism is best interpreted as admitting relational essences. That is, the correct characterisation of Putnam's natural kind essentialism is that which includes [3*], and not [3].

5. CONCLUSION

With Putnam's essentialism open to relational essences, it follows that traditional essentialism—the 'Kripke *and* Putnam' view—is free of (many of) the charges against it. The biological counterexamples to [3] thought to bring about the 'failure' and 'death' of essentialism no longer apply, and 'neo' essentialism turns out to be nothing new. One aspect of philosophical lore needs amending.

As a final word, I want to offer a brief comment on Kripke's natural kind essentialism. For the most part, I take the standard characterisation (that which includes [3]) to be the right interpretation of his views. On the rare occasions where he speaks of natural kind essences at all, he speaks of 'internal' structures, which should clearly be read as intrinsic: "Even though we don't *know* the internal structure of tigers, we suppose—and let us suppose we are right—that tigers form a certain species or natural kind."[28] However, I think its possible that a case could be made for a more liberal interpretation. I will not attempt that here, but I will briefly mention how it might go. In his account of the essentiality of origin, Kripke tells us that it is part of an individual person's essence that she comes from the particular sperm and egg she does. This is an essence of origin, and clearly relational. If natural kinds were to be treated as abstract individuals (as has been suggested for biological kinds), then a case could be made for their origin forming

[28] Kripke (1980: 120-121).

part of their essence. In which case Kripke's natural kinds could have relational essences. Food for thought.

REFERENCES

Achinstein, Peter (1968) *Concepts of Science – A Philosophical Analysis*. Baltimore: Johns Hopkins Press.

Boyd, Richard (1999) "Homeostasis, Species, and Higher Taxa". *Species – New Interdisciplinary Essays*. Wilson, R (Ed.). Cambridge: MIT Press, 141-185.

Ellis, Brian (2001). *Scientific Essentialism*. Cambridge University Press, New York.

Ellis, Brian. (1999). "Bigelow's Worries about Scientific Essentialism". in Howard Sankey, (Ed.), *Causation and the Laws of Nature*. Kluwer, Netherlands: 61-76.

Ereshefsky, Marc, (2007) "Species, Taxonomy, and Systematics". *The Handbook of Philosophy of Biology*, Matthen, Mohan and Stephens, Christopher (Eds.) Netherlands, Elsevier Publishing: 403-428.

Ereshefsky, Marc, and Matthen, Mohan (2005) "Taxonomy, polymorphism, and History: an Introduction to Population Structure Theory". *Philosophy of Science*, 72: 1-21.

Goosens, William (1977) "Underlying Trait Terms" *Naming, Necessity, and Natural Kinds*. Schwatrz, Stephen (Ed.). Ithaca: Cornell University Press: 133-154.

Griffiths, Paul (1999) "Squaring the Circle: Natural Kinds with Historical Essences" *Species – New Interdisciplinary Studies*. Wilson, R (Ed.). 1999. Cambridge: MIT Press, 209-228.

Hacking, Ian. (2007) "Putnam's Theory of Natural Kinds and Their Names is not the Same as Kripke's". *Principia*, 11: 1-24.

Hume, David. (1975) [1777] *Enquiries Concerning Human Understanding and Concerning the Principles of Morals* – 3rd Edition. Cambridge University Press. Republished by Dover, New York.

Kripke, Saul (1980) [1971] *Naming and Necessity*. Cambridge: Harvard University Press.

Lange, Marc *(forthcoming)* "The End of Diseases". *Philosophical Topics*.

Lange, Marc (2007) "Introduction to Part VI". *Philosophy of Science – An Anthology*. Malden, Blackwell Publishing: 371-377.

LaPorte, Joseph (2004) *Natural Kinds and Conceptual Change*. New York: Cambridge University Press.

Lewis, David (1986a) "Causation". *Philosophical Papers Vol. 2*. New York, Oxford University Press: 159-213.

Lewis, David (1986b) *On the Plurality of Worlds*. New York, Blackwell.

Lewis, David (1983) "Extrinsic Properties". *Philosophical Studies*, 44: 197-200.

Machery, Edouard (2005) "Concepts are not a Natural Kind". *Philosophy of Science*, 72: 444-467.

Mackie, Penelope (2006) *How Things Might Have Been*. New York: Oxford University Press.

Mellor, D H (1977) "Natural Kinds". *British Journal for the Philosophy of Science*, 28: 299-312.

Okasha, Samir (2002) "Darwinian Metaphysics: Species and the Question of Essentialism". *Synthese*, 131: 191-213.

Putnam, Hilary (1983) *Realism and Reason – Philosophical Papers Volume 3*. New York: Cambridge University Press.

Putnam, Hilary (1975) *Mind, Language and Reality – Philosophical Papers Volume 2*. New York: Cambridge University Press.
Salmon, Nathan (1982) *Reference and Essence*. Princeton: Princeton University Press.
Salmon, Nathan (1979) "How Not to Derive Essentialism from the Theory of Reference". *Journal of Philosophy*, 76: 703-725.
Slater, Matthew (2005) "Monism on the One Hand, Pluralism on the Other". *Philosophy of Science*, 72: 22-42.
Sober, Elliott (2000) *Philosophy of Biology – 2nd Edition*. Boulder, Colorado: Westview Press.
Sulmasy, Daniel (2005) "Disease and Natural Kinds". *Theoretical Medicine and Bioethics*, 26: 487-513.
Swain, Marshall (1980) "Causation and Distinct Events". *Time and Cause*. Van Inwagen, Peter (Ed.). 1980. Holland: Reidel: 155-169.
Thagard, Paul (1999) *How Scientists Explain Disease*. Princeton: Princeton University Press.
Whitbeck, Caroline (1977) "Causation in Medicine: The Disease Entity Model". *Philosophy of Science*. 44: 619-637.
Whitbeck, Caroline (1976) "The Relevance of the Philosophy of Medicine for the Philosophy of Science". *Proceedings of the PSA*. 44: 619-637.
Wilkerson, T.E. (1998) "Recent Work on Natural Kinds". *Philosophical Books*, 39: 225-233
Wilkerson, T.E. (1995) *Natural Kinds*. Brookfield, Vermont: Ashgate.
Wilkerson, T.E. (1986) "Natural Kinds and Identity, A Horticultural Inquiry". *Philosophical Studies*, 49: 63-69.
Wilson, Robert (1999) "Realism, Essence, and Kind: Resuscitating Species Essentialism?" *Species – New Interdisciplinary Studies*. Wilson, R (Ed.). 1999. Cambridge: MIT Press, 187-207.