What Externalism Is Not (Necessarily)

Tomohiro Sakai

t-sakai@waseda.jp

Abstract

Little has been written and much has been misunderstood by linguists about externalism. The purpose of this paper is to dispel some of the misunderstanding surrounding externalism. First of all, externalism does not hold that no mental entity is involved in understanding a word. Stereotypes are, or ought to be, in the mind, even though they are often insufficient, or even inappropriate, to determine the extension of the word in question. Secondly, externalism does not consider natural kind terms such as 'water' to belong to the category of context-dependent expressions on a par with absolutely indexical words such as 'I'. While the reference, but not the linguistic meaning, of 'I' depends on the context of utterance, the linguistic meaning of 'water' cannot be determined unless we take into account the physical environment in which the word form is tokened. Thirdly, externalism does not necessarily imply natural kind essentialism. Although Kripke is committed to such modal notions as 'possible world', 'rigid designation' and '(metaphysical) necessity', Putnam's Twin Earth thought experiment does not hang on any modal notions which may lead to the view that water = H_2O in all possible worlds. Finally, externalism does not hold that meaning and mental content are outside the head. Rather, it is the view that meaning and mental content are partly dependent for their individuation on one's external environment. Although linguistics has largely opted for internalism or individualism, no decisive argument has been adduced to demonstrate that everything about externalism is irrelevant to linguistics.

1. Externalism

For whatever reason, as Deutsch's (1993: 410) states, "[r]elatively little has been written by linguists on the Kripke-Putnam view", a view (or, more appropriately, views) commonly referred to as externalism. Externalism holds that meaning and mental content are partly dependent for their individuation on one's external environment. If externalism is true, then two speakers who are molecule-for-molecule identical may nevertheless have different thoughts and mean different things by employing exactly the same word forms. Externalism constitutes "current orthodoxy in the philosophy of mind" (Hunter 2003: 724) as well as in the philosophy of language (Häggqvist and Wikforss 2018: 912).

The trouble with the internalist conception of meaning and mental content is that intension or sense (Frege 1892/1997) is supposed to do too many things at once (Grabarczyk 2016: 155). Internalism holds, on the one hand, that intension (sense) is what a speaker understands when she understands meaning, and,

on the other hand, that intension (sense) is considered to determine extension (reference). In many cases, however, these two functions are not fulfilled simultaneously, because what a speaker understands about a word is often insufficient to determine its extension. Putnam (1973, 1975) invites us to imagine a planet, the Twin Earth, where our exact molecule-for-molecule replicas exist. Our replicas have exactly the same life history as we do, uttering the same word forms, having the same perceptual experiences, consuming the same food, and so on. The Earthian Oscar, for example, has a Doppelgänger, $Oscar_2$, on Twin Earth. Oscar speaks English, while $Oscar_2$ speaks Twin English, a language indistinguishable from our English. When Oscar feels pain in his leg, $Oscar_2$ does, too. When Oscar utters, 'My leg hurts', $Oscar_2$ does, too. Oscar and $Oscar_2$ are "exact duplicates in appearance, feelings, thoughts, interior monologue etc." (Putnam 1973: 702, 1975: 141)[1]. The physical environment of the Twin Earth is also indistinguishably similar to that of Earth in every respect. There are replicas of the Sahara desert, Mount Everest, the Pacific Ocean, the Nile River, and so on. The sole difference between the two planets is that on the latter the stuff called 'water' has a complicated molecular structure abbreviated as XYZ, instead of H_2O, although it is superficially indistinguishable from the Earthian water[2]. The Pacific Ocean and the Nile are filled with H_2O, whereas the Twin Pacific Ocean and the Twin Nile are filled with XYZ. On our planet, H_2O falls as rain, whereas on Twin Earth XYZ falls from the sky. The difference is not consciously accessible to the inhabitants on the two planets. Both Oscar and $Oscar_2$ associate with the stuff called 'water' descriptions such as 'liquid which is colorless, transparent, tasteless, thirst-quenching, found in lakes, etc.' (Putnam 1975: 191). Nevertheless, Putnam argues, the extension of the English word 'water' is H_2O, whereas that of the Twin English word 'water' (translated as 'twater' in our English; see Burge 1982: 100/2007: 85) is XYZ. Although both words are spelled as 'water' and pronounced as [ˈwɔːtə(r)], the Earthian word 'water' is not true of XYZ and the Twin-Earthian word 'water' is not true of H_2O (Segal 2000: 23). Kripke (1980) expresses a similar view:

> We identified water originally by its characteristic feel, appearance and perhaps taste, (though the taste may usually be due to the impurities). If there were a substance, even actually, which had a completely different atomic structure from that of water, but resembled water in these respects, would we say that some water wasn't H_2O? I think not. We would say instead that just as there is a fool's gold there could be a fool's water ; a substance which, though having the properties by which we originally identified water, would not in fact be water. And this, I think, applies not only to the actual world but even when we talk about counterfactual situations. If there had been a substance, which was a fool's water, it would then be fool's water and not water (Kripke 1980: 128)

The natural conclusion to draw is that "[t]he extension of our terms depends upon the actual nature of the particular things that serve as [local] paradigms, and this actual nature is not, in general, fully known to the speaker" (Putnam 1973: 711, 1975: 164). A bit surprisingly, this view entails that "the term 'water' had the

---

[1] This assumption is debatable. See below.

[2] Some may doubt that $Oscar_2$ is an exact replica of Oscar, since Oscar's body contains a significant amount of H_2O. As Stich (1978: 589, n. 16) and Farkas (2006: 326) remark, those who think that this fact is detrimental to the thought experiment are urged to construct a case in which the difference between the two planets concerns kinds of substances that are not found in the human body.

same extension in 1750 and in 1950" (Putnam 1973: 702, 1975: 141), since the actual nature of water remains the same all along. Before 1750, no one knew that water consisted of hydrogen and oxygen, and water was merely considered to be colorless, transparent, tasteless, thirst-quenching liquid found in lakes, etc. It is in the late eighteenth century that Henri Cavendish and others discovered that water was not an element but a compound (Barber ed. 2003: 25). What changed between 1750 and 1950 is not the meaning of 'water' but our understanding of the nature of water. The meaning of 'water' is determined by the nature of the physical environment, not by our understanding.

Now, we should be skeptical about Putnam's assumption that Oscar and Oscar$_2$ have exactly the same thought. Although Putnam's Twin Earth though experiment concerns not so much mental content as linguistic content (Noonan 1984: 216-217, Stalnaker 1993: 299-300, Woodfield 2000: 434, Liu 2002: 386, Sainsbury and Tye 2012: 4, n. 5), in most cases a difference in meaning yields a difference in thought. McGinn (1977: 531) states that "a correct specification of the mental states of the two groups of speakers in respect of H_2O and XYZ would mention those very substances", making it impossible to say that Oscar and Oscar$_2$ are in the same mental state. Suppose that Oscar and Oscar$_2$ utter (1):

(1)   There is some water within twenty miles, I hope. (Burge 1982: 101/2007: 86)

As Segal (2000: 24) remarks, "the content of the belief they express is just the content of the sentence they utter" (see also Stich 1978: 581). Oscar's thought corresponding to (1) is true if and only if there is water (H_2O) within twenty miles, whereas Oscar$_2$'s thought corresponding to (1) is true if and only if there is twater (XYZ) within twenty miles. If there is water but not twater around the two thinkers, only Oscar's thought corresponding to (1) is true. Since, in general, one and the same thought cannot be both true and false (Stich 1978: 578-579), it follows that Oscar's thought is different from Oscar$_2$'s, despite their internal identity[3]. If such is the case, Burge (1982: 102/2007: 87) claims, "the two are in no sense exact duplicates in their thoughts", as against Putnam's assumption. Neither meaning nor mental content can be described independently of relations the speaker/thinker bears to her environment (Burge 1993/2007: 300-301).

Burge (1979/2007) uses the word 'arthritis' to establish another form of externalism, namely social externalism, according to which meaning and mental content partly depend on the speaker's social environment, her internal state being insufficient for their individuation. Burge's (1979/2007) best-known thought experiment consists of three steps. First, suppose that Adam, a rational person generally competent in English, expresses his fear to his doctor by uttering (2):

(2)   I have arthritis in my thigh.

Adam has a number of beliefs about the disease called 'arthritis'. Thus, he thinks correctly "that he has had arthritis for years; that his arthritis in his wrists and fingers is more painful than his arthritis in his ankles, that it is better to have arthritis than cancer of the liver, that stiffening joints is a symptom of arthritis, that certain sorts of aches are characteristic of arthritis, that there are various kinds of arthritis, and so forth"

---

[3] We can also ask whether Oscar's and Oscar$_2$'s actions are the same, given that Oscar drinks H_2O while Oscar$_2$ drinks XYZ. The answer depends on how we individuate actions (Grabarczyk 2016: 166).

(Burge 1979: 77/2007: 104). Among such beliefs about arthritis is that expressed by the sentence in (2). The doctor tells Adam that this cannot be the case, because arthritis is specifically an inflammation of the joints. The patient relinquishes his view and now believes that some disease distinct from arthritis lodges in his thigh. As a second step, we are to conceive of a hypothetical community in which $Adam_2$, an exact replica of Adam, expresses his fear to his doctor by uttering (2). Up to this point exactly the same things happen to Adam and $Adam_2$. A major difference emerges when the doctor replies to $Adam_2$. In the hypothetical community, the word 'arthritis' is defined to apply to various rheumatoid ailments, including not only arthritis but also ailments afflicting a person's thigh. It follows that, unlike Adam's fear expressed by (2), $Adam_2$'s fear expressed by (2) is well grounded and may even be true. Finally, the counterfactual situation is interpreted. It is reasonable, according to Burge, to suppose that the community in which $Adam_2$ resides has no concept of *arthritis*, Adam's 'arthritis' and $Adam_2$'s 'arthritis' being two homonymous words which happen to share the same phonological form. Although the hypothetical community has a word spelled as 'arthritis' and pronounced as /ɑːθˈɹaɪtɪs/, it nevertheless lacks the concept of *arthritis*. $Adam_2$'s 'arthritis' denotes a disease distinct from arhtirits, say *tharthritis*[4]. On the standard conception of meaning, intension or concept uniquely determines extension (but not vice versa) and a difference in extension entails a difference in intension or concept. Therefore, as Burge (1979: 79/2007: 106) puts it, "[t]he word 'arthritis' in the counterfactual community does not mean *arthritis*", because it is not even extensionally equivalent to the actual word 'arthritis'. Since a thought consists of concepts, it also follows that Adam and $Adam_2$ have different thoughts about what they call 'arthritis'. Even though Adam and $Adam_2$ are molecularly identical, their thoughts differ, due to the sole fact that different meanings are associated with the same phonological form:

> The upshot of these reflections is that the patient's mental contents differ while his entire physical and non-intentional mental histories, considered in isolation from their social context, remain the same[5]. (We could have supposed that he dropped dead at the time he first expressed his fear to the doctor.) The differences seem to stem from differences "outside" the patient considered as an isolated physical organism, causal mechanism, or seat of consciousness. The difference in his mental contents is attributable to differences in his social environment. (Burge 1979: 79/2007: 106)

Meaning and mental content are not in the head but dependent for their individuation on one's social environment.

As Barber (ed.) (2003: 26) expounds, "[l]inguistics is no longer a branch of individualist psychology if it must consider an individual's physical and social environment before it can identify the content of the theory supposedly known by speakers". Nevertheless, linguistics has largely opted to go with individual psychology, as illustrated by Langacker's (2008) remark:

---

[4] A term coined by Burge (1979: 94/2007: 123).

[5] A mental state is said to be intentional when it is about objects or events in a world (McGeer 1994: 451, Rowlands et al. 2020: Section 3). For this reason, intentionality is often called 'aboutness'. Mental states which have intentionality or aboutness such as beliefs, desires, fears and hopes are called propositional attitudes (cf. Brown 2004: 2).

> Our concern is with the meanings of linguistic expressions. Where are these meanings to be found? From a cognitive linguistic perspective, the answer is evident: meanings are in the minds of the speakers who produce and understand the expressions. It is hard to imagine where else they might be. Yet there are many scholars who resist or reject that answer. (Langacker 2008: 27)

One of the reasons why linguistics has expressed little interest in externalism is that externalism is primarily couched in terms of the notion of 'truth' (Sakai 2022a: 26), a notion related not only to 'truth condition' (Burge 1982: 110, 115/2007: 93, 97, Haukioja 2017: 869, Wikforss 2008: 162, 169, 179) and 'truth value' (Burge 1979: 99/2007: 128), but also to 'extension' (Quine 1951: 21, Putnam 1975: 134, 154) and 'reference' (Kripke 1980: 135-136, Hacking 2007a: 3). Truth and extension are inseparable insofar as "[t]he extension of a term is just what the term is *true of*" (Putnam 1975: 154, emphasis in the original). Such truth-related notions are external to the language faculty of competent language users (Chomsky 1955: 40-41), making externalism seemingly irrelevant to linguistics.

This does not establish, however, that everything about externalism is extraneous to linguistics. For Langacker (2008: 4), grammar is "an integral part of cognition" and "reflects our basic experience of moving, perceiving, and acting on the world". According to this conception of language, the study of the relationship between semantic content and environment constitutes part and parcel of genuine linguistic inquiry, on which the internalism vs. externalism debate may have a direct bearing. The purpose of the subsequent argument is not to settle the internalist-externalist debate, but merely to dispel some of the misunderstanding surrounding externalism. A full comparison between externalism and internalism would take me too far afield and has to await another occasion. For the same reason, I shall not go into the differences between physical externalism (Putnam 1973, 1975) and social externalism (Burge 1979/2007)[6].

2.   No Mental Entity Involved?

Externalism does not hold that no mental entity is involved in understanding a word (Wikforss 2008: 178, n. 2). Putnam (1973, 1975) and Kripke (1980) put forward the view that the meaning of the word 'water' is not solely defined in terms of the description of the superficial properties of water and that the underlying nature, possibly unknown, of the local paradigm of water (Donnellan 1993: 156) must be taken into account, making Oscar's and Oscar$_2$'s uses of the same word form 'water' occurrences of two distinct words, even though the two speakers are molecule-for-molecule identical. This view does not entail, however, that extension or reference exhausts word meaning. Even though Kripke emphasized that the meaning of natural kind terms cannot be equated with any description, as Soames (2002: 241) states, "he did not provide a positive doctrine that identifies what the semantic contents" of such terms are[7]. On Häggqvist's and Wikforss's (2020: 225) interpretation, Kripke held the view that the semantic content of a natural kind term "is exhausted by the term's extension (across possible worlds), just like reference is held to exhaust the semantic content of proper names". This may give the impression that externalism is entirely

---

[6] See Liu (2002) for some discussion.
[7] Soames here talks about the semantics of proper names, but the same point applies to natural kind terms, since Kripke (1980) aimed to extend his view about proper names to a wider class of terms including natural kind terms (Soames 2002: 242, Wikforss 2013: 243).

what Lakoff (1987: 157ff) refers to as "objectivist" semantics which roughly consists in equating meaning with extension or reference. But this is a misrepresentation of externalism. In fact, as Lakoff (1987: 169) notes, Putnam's (1973, 1975) view is presented "in a somewhat different form" from Kripke's (1980) and, as Hacking (2007a: 3) points out, "unlike Kripke, Putnam did offer an explicit meaning of 'meaning'". Lakoff (1987: 169) acknowledges that "Putnam's […] account of meaning is objectivist in most, but not quite all, ways". The meaning of a word includes, over and above its reference, what Putnam called its stereotype, that is, in Hacking's (2007a: 3) words, "some sort of common (if possibly mistaken) knowledge associated with the [word]"[8]. Thus, Putnam states that "speakers are required to know something about (stereotypical) tigers in order to count as having acquired the word 'tiger'; something about elm trees (or, at least, about the stereotype thereof) to count as having acquired the word 'elm'; etc." (Putnam 1975: 168). On Putnam's view, "'meaning' never means 'extension'" (Putnam 1975: 140), even though "[t]he extension is included in the meaning" (Hacking 2007a: 3)[9]. In this respect, as Abbott (1997: 313) remarks, "Putnam's theory is less radical than Kripke's view". Assuming that the extension is included in the meaning, a distinction must be made between acquiring a word and learning its meaning (Putnam 1975: 167). One can acquire a word, that is, store a word in one's mental lexicon, without learning its meaning.

The minimum level of knowledge necessary for a speaker to be considered to have acquired a word differs from word to word and from culture to culture. Speakers of English are required to be able to tell tigers from leopards, but not required to be able to tell elm trees from beech trees (Putnam 1975: 168). As Recanati (1993: 148) observes, the learning of word meaning is a matter of degree. At one end of the spectrum there are words such as 'phoneme' which are unknown to average users of the language. There are also words whose existence, but not whose stereotype, is known to average speakers, such as 'molybdenum'. In still other cases, the stereotype associated with words is very poor, as is the case with words like 'elm' and 'beech', both of which are associated with the stereotype "common sort of deciduous tree" (Putnam 1992: 386, see also Putnam 1975: 143, 147). At the other end of the spectrum, we find words such as 'water' with which average speakers associate a fairly rich stereotype. Speakers not familiar with the stereotype of a word required by the community are not deemed competent in the use of the word. For instance, as Putnam (1975: 168) states, there is no point in talking about tigers with a person who, pointing to snowball, asks, 'Is that a tiger?' (Putnam 1975: 168). The right thing to say in this case is that the person fails to refer with the utterance of 'tiger'. The Twin-Earth thought experiment rests on the assumption that both Oscar and Oscar$_2$ are familiar with the stereotype associated with the word form 'water'; otherwise there would be reference failure when they employed 'water', and no thesis about the public meaning of

---

[8] According to Putnam's own definition, a stereotype is "a description of what a typical speaker thinks a paradigmatic 'elm', or whatever, is (or is conventionally assumed to be) like" (Putnam 1992: 386). One of the examples of radically mistaken stereotypes, Putnam (1975: 170) says, is that associated with the word 'witch', namely "witches enter into pacts with Satan", "they cause sickness and death", and the like. Even though there exist such mistaken stereotypes, "[m]ost stereotypes do in fact capture features possessed by paradigmatic members of the class in question" (Putnam 1975: 170).

[9] Not every externalist shares this view. Bilgrami (1992: 2), for instance, states that one can "be an externalist and yet deny that reference in *any* standard sense has anything to do with content". This version of externalism will not be the primary object of the present discussion, because, as Putnam (1992: 386-397) makes clear, it has little to do with linguistic semantics:

> But Bilgrami's interests are quite different. In his recent *Belief and Meaning* [= Bilgrami (1992)] he makes it clear that his purpose is not to explicate the concept of meaning at all; indeed, he argues that for the purpose of psychological explanation, what we want is *not*, in general, a knowledge of the meaning of the thinker's words. *Meaning* is not all that important a notion, he thinks. But be that as it may, it is the notion I was trying to explicate in MoM [= Putnam (1975)]! (Putnam 1992: 386-387)

'water' would follow. The gist of Putnam's (1975) argument is that, although stereotypes are, or ought to be, in the mind, they are often insufficient, or even inappropriate, to determine the extension of the word in question and the content of the thought the subject entertains by using that word. In most cases, the stereotype associated with a word constitutes neither its intension nor its sense. Fodor (1975: 61) summarizes that, if Putnam is right, "either the semantic properties of a word aren't what you learn when you learn the word, or the semantic properties of a word don't determine its extension". This summary is correct provided that the first occurrence of 'semantic properties' is intended to mean 'extension' and the second 'stereotype'[10].

Needham (2011: 2-4) maintains that the use of 'etc.' in the stereotypical description of 'water' offered by Putnam (1975: 191), 'liquid which is colorless, transparent, tasteless, thirst-quenching, found in lakes, etc.', is problematic in that it makes room for the possibility that the description may be completed into the necessary and sufficient condition for something to be water. This would, Needham argues, undermine Putnam's and Kripke's claim that the meaning of 'water' can only be defined in terms of the microstructure of water, because the completion of the stereotypical description may be effected solely by appeal to the macroscopic properties of the stuff. This argument may be adequate if it is directed at microessentialism, according to which a chemical substance can only be defined by its microstructure. However, the use of 'etc.' in the stereotypical descriptions of 'water' undermines neither the notion of stereotype nor the externalist thesis that what is in the head is not sufficient to determine the meaning of a word. After all, the use of 'etc.' indicates that we are not required to fill out the 'etc.' in order to be competent in the use of the word 'water'. We must know the stereotype associated with 'water', but the stereotype does not have to be specific enough to determine the meaning of the term.

3.   Indexicality?

The second point to be made about externalism, which is the reverse of the first point just rehearsed, is that the externalist position does not consist in reducing the meaning of every word to an indexical meaning exhibited by a pure indexical like 'I', a meaning which is generally assumed to reside in the head. Pure indexicals are words whose content varies from context to context. The pronoun 'I', for instance, refers to different people depending on the context of use; 'I' refers to Mary when uttered by Mary, and refers to John when uttered by John. Nevertheless, despite the variation of its content, 'I' always means 'the utterer', independent of the context of use (Benveniste 1956/1966: 252, Recanati 1993: 4.3). Kaplan (1989: 505ff) calls the constant linguistic meaning of indexicals 'character'. Competent speakers of English know the character of 'I', which enables them to pick out a particular individual in each context. As regards indexicals, then, we can say that what is in the mind, i.e. the character, fails to determine the extension[11], the extension being determined only with respect to particular contexts (Burge 1979: 86/2007: 114-115). This does not mean that what is in the mind fails to determine the meaning, since the character is in the mind. The linguistic meaning of a pure indexical word like 'I' is its character, not the object picked out in a given context. This reasoning is not applicable to natural kind terms, because, as Burge (1982: 103/2007:

---

[10] To be more faithful to Putnam's (1975: 167) terminology, we should say that the semantic properties of a word aren't what you learn when you *acquire* the word.
[11] Kaplan (1989) calls the reference of an indexical expression 'content', rather than 'extension'.

88) emphasizes, "it is clear that 'water', *interpreted as it is in English*, or as we English speakers standardly interpret it, does not shift extension from context to context in this way" (see also Wikforss 2008: 169). Even if Oscar utters (1) above on Twin Earth, the reference of 'water' is still water (H_2O) rather than twater (XYZ) (Farkas 2006: 333)[12], and his utterance is false, since there is no water on Twin Earth. The extension of 'water' remains constant across contexts, unlike pure indexicals like 'I' or 'here' (Putnam 165, Bach 1987: 263, Burge 1989: 181/2007: 283).

What is puzzling, as Bach (1987: 263) puts it, is that Putnam states in several places that 'water' is indexical (Putnam 1973: 710-711, 1975: 152, 162, 187, 188, 193). Following Putnam, De Brabanter and Leclercq (2019: 4ff) employ the term 'indexical externalism' (externalisme indexical) to refer to physical externalism. But if 'water' is indexical as 'I' is, one could argue, contrary to externalism, that the meaning (character) of 'water' is in the mind just as the meaning (character) of 'I' is, and that 'water' has the same meaning on Earth and on Twin Earth (Putnam 1975: 164-165). This is the reason that Burge (1982/2007) objects to the idea that natural kind terms are indexicals. However, when Putnam (1975) claims that natural kind terms are indexicals, he does not mean that 'water' has a context-independent character as 'I' does (Putnam 1975: 165). As Liu (2002: 400, n. 3) points out, "Putnam is using 'indexical' in a way different from the standard usage of the term". What Putnam arguably had in mind is the fact that the reference of 'water' cannot be determined unless we take into account local paradigms of 'water', which are only identified by way of ostensive definitions such as "'water' means 'whatever is like water, bears some equivalence relation, say the same liquid relation to *our* water'" (Putnam 1974: 451) or "this liquid is called water" (Putnam 1973: 702, 1975: 141). As Wikforss (2008: 160) states, "indexical elements play a role when these terms are introduced into the language". Putnam's notion of indexicality alludes to the fact that the local paradigms of a term can only be fixed through direct reference (Liu 2002: 399). The indexical elements 'our' and 'this' occurring in the definition of the word 'water' refer to the environment in which 'water' is tokened (Linsky 1977: 823). This enables Oscar's and Oscar$_2$'s uses of 'water' to mean different kinds of stuff, despite their internal identity. While the meaning of 'water' *per se* is non-indexical, the determination of the meaning of 'water' necessarily involves an indexical component by which a particular kind of stuff is picked out on each planet and without which Oscar's and Oscar$_2$'s uses of 'water are indistinguishable (Donnellan 1993: 159). As Donnellan put it, "the importance of indexicality in the Putnam thought experiments really shows up at a deeper level, at the level of "explaining the meaning of natural kind term". Linsky (1977: 822) remarks that natural kind terms and indexicals have in common that the extension "is not determined by any properties it expresses but rather by the situation of its utterance".

---

[12] If Oscar is transported to Twin Earth without his knowledge and stays there long enough to think constantly about the watery stuff before him, it can happen that the reference of his use of 'water' shifts to twater (XYZ) (Brown 2004: 39-40, Farkas 2006: 337, Sainsbury and Tye 2012: 92-93). Burge calls this shift 'slow switching':

[…] if one were stealthily shifted back and forth between actual situations that modeled the counterfactual situations, one would not notice some feature in the world or in one's consciousness which would tell one whether one was in the "home" or the "foreign" situation. […] The thoughts would switch only if one remained long enough in the other situation to establish environmental relations necessary for new thoughts. So quick switching would not be a case in which thoughts switched but the introspection remained the same. But slow switching could be such a case. (Burge 1988: 652)

As Grabarczyk (2016: 157) states, "the longer the user uses the term in contact with XYZ the more the context shifts towards Twin Earth". Pure indexicals, by contrast, do not need an interval to change their reference. For instance, Oscar's use of 'this liquid' on Twin Earth immediately picks out twater.

This is supposedly what Putnam (1973: 710, 1975: 152) meant by saying that "words like 'water' have an unnoticed indexical component". In other respects, natural kind terms such as 'water', 'tiger' and 'lemon' bear scant resemblance to indexicals such as 'that', 'now' and 'today' (Wiggins 1995: 61-62). Indeed, Putnam drew a distinction between natural kind terms and "obviously indexical" or "absolutely indexical" words like 'I' and 'here' (Putnam 1973: 710, 1974: 148, 1975: 152, 165, Wikforss 2008: 160). As a consequence, Burge (1982: 103/2007: 88) is right to say that "[t]here is nothing at all indexical about 'water' in the customary sense of 'indexical'", as long as 'customary' indexicals are identified with what Putnam called 'absolutely' indexicals.

4. Essentialism?

Externalism does not necessarily force us to embrace natural kind essentialism, according to which natural kinds such as water, cat or gold have essential properties, namely properties instantiated in all possible worlds. Natural kind essentialism holds for example that "in any possible world anything that is not an animal is not a cat and that in any possible world anything that is not composed of molecules of H_2O is not water" (Robertson Ishii and Atkins 2020: Section 4). On this view, being H_2O is the essential property of water, while being transparent, thirst-quenching, etc. is merely an accidental property. Essentialism involves such modal notions as 'possible world', 'rigid designation' and '(metaphysical) necessity'. Some externalists are committed to essentialism in this sense, giving the impression that 'externalism' and 'essentialism' are two names of the same doctrine. Thus, Kripke (1980) talks about necessary truths about natural kinds such as cow, tiger, water and gold, defined in terms of the underlying essential properties that such kinds have in all possible worlds, in some of which the actual physical laws even do not hold. For Kripke, the statements in (3) express necessary or metaphysical truths, that is, truths which hold in all possible worlds (cf. Soames 2002: 244):

(3)   a.   Water is H_2O.
        b.   Flashes of lightening are flashes of electricity.
        c.   Cats are animals.
        d.   Whales are mammals.
        e.   Heat is the motion of molecules.

Necessary statements must be distinguished from analytic statements, namely statements that are true solely by virtue of meaning. The analytic statements in (4) are *a priori* in that their validity can be confirmed without any empirical investigation:

(4)   a.   All bachelors are unmarried.   (Putnam 1962: 366)
        b.   All vixens are foxes.   (Putnam 1962: 366)

Necessary statements, by contrast, may be *a posteriori* in that they are only confirmed empirically. Kripke (1980: 128) states that "science can discover empirically that certain properties are *necessary* of cows, or of tigers". Once confirmed in the actual world, Kripke argues, the statements in (3) are necessarily true,

because, like proper names, natural kind terms are rigid designators, i.e. terms that designate the same object or kind in all possible worlds[13]. If water = H_2O in the actual world and both 'water' and 'H_2O' are rigid designators, then there is no world in which the two terms designate different things. Therefore, it is true in all possible worlds that water = H_2O. Kripke's view resembles Putnam's in dictating that superficial properties of a substance may fail to define the substance:

> Let us suppose the scientists have investigated the nature of gold and have found that it is part of the very nature of this substance, so to speak, that it have the atomic number 79. Suppose we now find some other yellow metal, or some other yellow thing, with all the properties by which we originally identified gold, and many of the additional ones that we have discovered later. An example of one with many of the initial properties is iron pyrites, 'fool's gold.' As I have said, we wouldn't say that this substance is gold. (Kripke 1980: 124)

The distinction between gold and iron pyrites (fool's gold) is analogues to that between water and twater, the sole difference being that twater but not iron pyrites is an imaginary substance[14]. Mellor (1977: 301) characterizes both Putnam's and Kripke's views as anti-Fregean by saying that "Putnam's theory of the extension of kind terms, and Kripke's theory of their reference, are alike in denying traditional accounts that make the reference (or extension) of terms a function *inter alia* of something like their Fregean sense". Donnellan (1983/2012: 179) goes so far as to suggest that there is a logical relation between Putnam's and Kripke's views: "[t]he theory of natural kinds terms developed by Saul Kripke and Hilary Putnam is seen by both authors, I believe, as being intimately connected to Kripke's views about reference, perhaps even a consequence of them". Based on the putative similarity, the two authors' views are often unified into the 'Kripke-Putnam' view or 'Putnam-Kripke' view (Hacking 2007a: 1-2, Wikforss 2013: 242). For example, Donnellan's (1983/2012: 180) discusses "the Kripke-Putnam treatment of natural kind terms" and Lakoff (1987: 173) regards "the Putnam-Kripke view" as "objectivist". Given Kripke's (1980) emphasis on necessary truths, one may naturally suppose that endorsing externalism is endorsing essentialism:

> So if this consideration is right, it tends to show that such statements representing scientific discoveries about what this stuff *is* are not contingent truths but necessary truths in the strictest possible sense. It's not just that it's a scientific law, but of course we can imagine a world in which it would fail. Any world in which we imagine a substance which does not have these properties is a

---

[13] Kripke (1980: 122, 127, 134) stresses the similarity between proper names and natural kind terms:
We use terms such as 'gold' and 'tiger' rigidly, as applying to 'that kind of thing'; consequently, the meaning of a natural kind term cannot be given a descriptivist construal any more than the meaning of a proper name. (Kripke, 1980, p. 122) According to the view I advocate, then, terms for natural kinds are much closer to proper names than is ordinarily supposed. The old term 'common name' is thus quite appropriate for predicates marking out species or natural kinds, such as 'cow' or 'tiger'. My considerations apply also, however, to certain mass terms for natural kinds, such as 'gold', 'water', and the like. (Kripke 1980: 127)
Kripke's doctrine of proper names, however, cannot be so straightforwardly extended to other terms, as discussed by Soames (2002) and Wikforss (2013), among others.
[14] Putnam also gives examples of actual kinds. Normal speakers of English associate with both 'elm' and 'beech' the stereotype 'common deciduous trees' (Putnam 1975: 143, 147), which does not even allow us to determine the local paradigms. In fact, gold and iron pyrites are not good examples of twin-substances, because, as Ben-Yami (2001: 161) points out, "iron pyrites has only a *faint* resemblance to gold".

world in which we imagine a substance which is not gold, provided these properties form the basis of what the substance is. In particular, then, present scientific theory is such that it is part of the nature of gold as we have it to be an element with atomic number 79. It will therefore be necessary and not contingent that gold be an element with atomic number 79. (Kripke 1980: 125)

This may appear to provide a sufficient reason for considering Putnam an essentialist, as Mellor (1977: 299) does: "Kripke and Putnam claim that natural kinds have essential properties; that is, properties which nothing can lack and still be of the kind". Kripke (1980) and Putnam (1973, 1975) certainly share the idea that the extension of a term is included as part of its meaning, the descriptions attached to the term being often insufficient to determine its extension, contrary to Frege's (1892/1997) idea that the sense of a term grasped by a speaker determines its reference. As Hacking (2007a: 11) points out, however, "Putnam seldom dabbled in essences". Hacking (2007a: 11-12, 2007b: 228) observes that Putnam almost always puts 'essence' and 'essential nature' in scare-quotes, presumably "indicating amusement or irony" (ibid.: 12). This subtle difference between Kripke and Putnam stems from their attitudes toward the notion of rigid designation. This notion is, Hacking (2007a: 5) states, "integral to Kripke's approach", while it is "at most incidental to Putnam's". Putnam (1992) is very clear about the difference between 'water' and 'H_2O'[15]:

> […] competence in the use of "H_2O" presupposes competence in the use of 'H' (hydrogen) and 'O' (oxygen), as well as in the use of the notions 'molecule' and 'atom'. The very different form of entry for 'water' makes it clear that the ability to characterize the extension of 'water' in some other way than simply by using the word 'water' (e.g., by a scientific definition) is not presupposed by competence in the use of the word (notwithstanding the fact that the extension of 'water', in one sense of that word, is 'H_2O give or take certain impurities' […] (Putnam 1992: 388)

Even though 'water' and 'H_2O' have the same extension, the two terms in no way express the same concept or intension (Burge 1979: 76/2007: 103). To say simply that both are rigid designators does not enable us to account for the fact that one can be competent in the use of 'water' without yet being competent in the use of 'H_2O'. A similar point is made by Fodor (1994: 57, 105-106), who claims that "despite their synonymy, the conditions for *having* the concepts [of water and H_2O] are different" (Fodor 1994: 57). This perspective is orthogonal to the simple idea that the meaning of a word can be defined in purely scientific terms, as illustrated by Bloomfield's (1933: 239) remark: "We can define the names of minerals, for example, in terms of chemistry and mineralogy, as when we say that the ordinary meaning of the English word *salt* is 'sodium chloride (NaCl)'".

Putnam goes further by saying that 'water' and 'H_2O' differ not only in intension but also in extension, because the extension of the latter "is the set of all substances that consist entirely of molecules consisting of two hydrogen atoms and one oxygen atom", to which most water does not belong (Putnam 1992: 406, n. 62). For Putnam, unlike for Fodor, the two terms are by no means synonymous. As noted by Gasparri and Marconi (2021: Section 3) in this connection, it is in principle possible to suppose that our chemistry

---

[15] Putnam (1992: 388) also says that, unlike 'water', 'H_2O' is a one-criterion term (cf. Putnam 1962).

is badly mistaken in assuming that water is H_2O. Importantly, however, even if water turned out not be H_2O, Putnam's claim that the meaning of 'water' depends on the physical environment in which the word is tokened would not be undermined, insofar as there are some important properties which, unbeknownst to competent speakers, distinguish water from twater. For Putnam, water is "whatever is of the same character, whatever has the same important physical properties, as *our* water" (Putnam 1974: 452). As Häggqvist and Wikforss (2018: 914) point out, in Putnam's argument, 'H_2O' is "a mere placeholder", albeit "a bit less explicitly so than XYZ"[16]. In a similar vein, Burge (2003: 457) claims that the force of Putnam's Twin Earth thought experiment "does not depend on the assumption that water is H_2O". It is not essential whether the important physical properties of water involve hydrogen and oxygen:

> 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. (Putnam 1975: 157)

Interest in 'interest' characterizes Putnam's position. As Hacking (2007a: 9) puts it, Kripke's and Putnam's views can be kept apart with two words: "essence for Kripke, and interest for Putnam". Putnam's non-essentialist position is clearly articulated in the following passage:

> […] we discover 'tigers' on Mars. That is, they look just like tigers, but they have a silicon-based chemistry instead of a carbon, based chemistry. (A remarkable example of parallel evolution!) Are Martian 'tigers' tigers? It depends on the context. (Putnam 1975: 157-158)

Putnam's answer to the question "Are Martian 'tigers' tigers?" sharply contrasts with Kripke's (1980: 121), according to which "anything not of this species, even though it looks like a tiger, is not in fact a tiger". If Martian tigers can be tigers, one may naturally suppose that twater can be considered a kind of water in contexts where the superficial characteristics of substances are more important than their chemical structures. This is indeed Putnam's position:

> And structure may sometimes be unimportant; thus one may sometimes refer to XYZ as water if one is using it as water. (Putnam 1975: 157)

All that is needed for the Twin Earth thought experiment is the assumption that there may be contexts in which twater is not water and that the difference between water and twater is possibly inaccessible to competent speakers. We do not have to appeal to modal notions such as 'possible world', 'rigid designation' and '(metaphysical) necessity'.

---

[16] Häggqvist and Wikforss (2018) do not endorse Putnam's externalism, even if 'H_2O' is interpreted as a mere placeholder. Rather, they rebel against Putnam's and Kripke's view on natural kind terms, concluding that "[t]ime is ripe to cut the cord with the legacy of Kripke and Putnam, and start afresh" (ibid.: 929).

What complicates the matter is that, as pointed out by Hacking (2007a: 5) and Wikforss (2013: 250), in a section entitled 'Indexicality and Rigidity', Putnam (1975: 146-152) seems to fully endorse Kripke's natural kind essentialism:

> Kripke calls a designator 'rigid' (in a given sentence) if (in that sentence) it refers to the same individual in every possible world in which the designator designates. If we extend the notion of rigidity to substance names, then we may express Kripke's theory and mine by saying that the term 'water' is *rigid*. (Putnam 1975: 148):

> Suppose, now, that I discover the microstructure of water – that water is H_2O. At this point I will be able to say that the stuff on Twin Earth that I earlier *mistook* for water isn't really water. In the same way, if you describe not another planet in the actual universe, but another possible universe in which there is stuff with the chemical formula XYZ which passes the 'operational test' for water, we shall have to say that that stuff isn't water but merely XYZ. You will not have described a possible world in which 'water is XYZ,' but merely a possible world in which there are lakes of XYZ, people drink XYZ (and not water), or whatever. In fact, once we have discovered the nature of water, nothing counts as a possible world in which water doesn't have that nature. Once we have discovered that water (in the actual world) is H_2O, *nothing counts as a possible world in which water isn't H_2O*. In particular, if a 'logically possible' statement is one that holds in some 'logically possible world,' it isn't logically possible that water isn't H_2O. (Putnam 1975: 150): "

Why was Putnam (1975: 146-152) so sympathetic toward Kripke's modal claim, on pain of misrepresenting his own view? According to Wikforss (2013: 252), this has to do with Putnam's distrust of the traditional positivist account of meaning and necessity, famously challenged by Quine (1951)[17]. For logical positivism, necessary statements are nothing but analytic statements, which are true by virtue of their meaning. If the statements in (3) are analytic, it must be the case that the word 'water' underwent a semantic change when it was discovered that water is H_2O. Before the discovery, 'water' was defined by the superficial characteristics of water, but now it is defined by the chemical structure of water. On Putnam's (1975: 142, 153, 1990: 60) account, on the contrary, the meaning of 'water' remains constant even after the discovery that water is H_2O. This suggests that, unlike the statements in (4), the statements in (3) are not analytic, hence not *a priori*. Putnam thought that the separation of necessity from analyticity and *a priori* was in line with Kripke's theory, according to which the statements in (3) are necessary and metaphysical but are neither analytic nor *a priori*. As indicated above, however, Kripke's view entails essentialism, with which Putnam's emphasis on interest is clearly incompatible.

Whatever the reason may be for Putnam's sympathy toward Kripke's modal claim, Hacking (2007a: 16) maintains that "[a]lthough in the pages just cited, Putnam heartily endorsed possible worlds, rigid designation and necessity, those pages can be deleted without affecting any position to which Putnam was himself later committed". Hacking's interpretation accords with Putnam's (1990) remark:

---

[17] For Quine's (1951) argument, see Sakai (2022b: 3.2).

> I do not think that a criterion of substance-identity that handles Twin Earth cases will extend handily to 'possible worlds'. In particular, what if a hypothetical 'world' *obeys different laws* ? […] I now think that the question, 'What is the necessary and sufficient condition for being water *in all possible worlds*?'makes no sense at all. And this means that I now reject 'metaphysical necessity'. I won't insist (any more) that 'it is conceivable that water may turn out not to be H_2O but it isn't logically possible that water isn't H_2O'. (Putnam, 1990: 69-70)

Externalism as such makes no commitment to reducing linguistic semantics to essentialism, which holds that chemical substances exhibit some invariable microscopic properties, namely essences, in all possible worlds[18].

5. Outside the Head?

The last, and probably most important point to be made about externalism is that, contrary to what the name might suggest, it is not a view about the locus of meaning and thought; at stake is not whether meaning and thought are inside or outside the head (Burge 2003: 454-455), but whether a person's intrinsic properties are sufficiently rich to individuate what the person means by his words vis-à-vis what his twin means by his words. It is therefore possible, for example, to be a Platonist about meaning and thought without being an externalist in the relevant sense. Frege (1918-1919: 69/1956: 302) famously (or infamously) considered senses and thoughts to belong neither to the inner world nor to the outer world of material, perceptible things, but to what he called the third realm ('drittes Reich'). For Frege, meanings are not mental entities but abstract entities found outside the mind of the thinker (Putnam 1973: 699, 1975: 134). This does not make Frege an externalist in the relevant sense, however, because he undoubtedly equated grasping or apprehending the sense of an expression with being in a certain state of mind; nowhere did he refer to the thinker's external environment as a factor contributing to the determination of her thoughts [19]. Although 'internalism vs. externalism' and 'individualism vs. anti-individualism' are "approximately interchangeable" and "[u]sage here is obviously a matter of taste" (Burge 2007: 154), Burge prefers the terms 'individualism' and 'anti-individualism' chiefly because the terms 'internalism' and 'externalism' would wrongly suggest that "the main issue is essentially concerned with spatial location" (ibid.). What is at issue is not spatial location but "the role of the individual and the individual's relations to a wider order" and the social environment is "a prominent subclass" of such relations (Burge 2007: 155).

---

[18] This is not to say that Putnam's view is not without its problems. Needham (2011) remarks:
> What could Putnam's twin earth fantasy show beyond the triviality that sufficiently ignorant people might not be able to distinguish similar substances? If it assumes that two substances are distinct at the microlevel and yet share all their macroproperties, so that they can't be distinguished in terms of macroproperties, then the scenario is wildly implausible. Assuming that it is in some sense possible doesn't show it to be possible. (Needham 2011: 11)

See Häggqvist and Wikforss (2018, 2020) for some discussion.

[19] Frege did refer to the context of utterance as a factor contributing to the expression of a thought (Frege 1918-1919: 64/1956: 296). Thus, 'I am hungry' uttered by John and the same sound uttered by Mary express different thoughts. But this merely suggests that Frege was aware of the existence of what Kaplan (1989) and Putnam (1973, 1975) called (absolutely) indexical words' as opposed to natural kind terms like 'water'. At stake in the internalist vs. externalist debate is whether the meaning of a non-indexical term can be determined in total independence of the nature of the environment in which it is tokened. Both parties accept the obvious fact that the reference of an absolutely indexical term can only be determined with respect to the context of utterance.

Externalism is not the view that meaning and mental content are outside the head, but the view that their individuation partly depends on "relations that are not reducible to matters that concern the individual alone" (ibid.: 154). Rowlands et al. (2020) make this point clear:

> Despite some initially over-exuberant formulations of Putnam (1975) – of the "Cut the pie any way you like, meaning just ain't in the head", variety – the correct conclusion of this thought experiment does not concern the *location* of (some) mental states but, rather, a claim about their *individuation*. External individuation does not entail external location. Sunburn, as Davidson (1987) pointed out, is externally individuated: individuation dependent on factors that lie outside the skin. Nevertheless, it is still located on the skin. That some mental states are individuation dependent on circumstances that exist outside the skin of an individual does not entail that those mental states are located outside the individual's skin. (Rowlands et al. 2020: Section 3.1, emphases in the original)

The question, "Are meanings in the head?", posed by Putnam (1973: 700, 1975: 139) must be interpreted rhetorically (Nuccetelli ed. 2003: 3). The internalist-externalist debate concerns whether or not the individuation of meaning and thought can be effected without taking into account the nature of the environment of which the utterer or thinker may be unaware.

It may be conceivable that, in saying that meaning is in the head, Langacker (2008: 27) is mistaken in taking literally the question, "Are meanings in the head?" (Sakai 2022b: 28). As Nuccetelli (ed.) (2003: 3) points out, the inside-outside distinction is only a metaphor. Literally speaking, it is evident that all mental states are in the head. Even Burge (1993: 453) agrees with this, stating that externalism, or what he calls anti-individualism, "is about the nature of 'internal' psychological state". To say that meaning is in the head in the literal sense of the term is not sufficient to demonstrate that meaning is in the head in the metaphorical sense of the term. The former claim amounts to a truism endorsed by both the internalist and the externalist, while the latter claim is incompatible with internalism[20]. It is more than obvious that Oscar's mental state is in his mind and that Oscar$_2$'s mental state is in his mind, if 'in' is interpreted literally. It would be ludicrous to claim that their mental states were found outside their mind, say, in a lake or in a conference hall. Far from repeating such a truism, the externalist holds that what distinguishes Oscar's and Oscar$_2$'s mental states are not to be sought for in their mind, claiming that "the existence and nature of certain psychological kinds depends necessarily on the existence and nature of certain relations to specific kinds or situations in the environment" (Burge 2003: 454). It appears to be difficult to distinguish the externalist thesis thus formulated from Langacker's (2008: 28-29) remark, made from the internalist perspective, that "[t]he conceptualizations we entertain are undeniably internal, in the sense of taking place in the brain, yet reach beyond it in the sense of being conceptualizations **of** some facet of the world"

---

[20] This point should be made more carefully. Frege (1918-1919/1956) held that sense and thought were in the third realm distinct from both physical and mental realms, while grasping the sense or thought is an individual mental act. If we accept this idea, along with the view that the Fregean sense is an aspect of the meaning of an expression, it will be more accurate to say that it is a truism that the mental process of understanding the meaning of an expression is in the individual's mind. Burge (2003) accepts this truism:

> Meaning is abstract and hence not anywhere. But the psychological state of understanding a meaning is naturally seen "in" the mind or brain. Nothing anti-individualism requires rejecting this natural view. I accept it (Burge 1982). (Burge 2003: 455)

Sakai, Tomohiro

(emphasis in the original). One may ask here whether it is possible to count Langacker among the externalists, by construing his apparently internalist claim that meanings are in the speaker's mind in such a way that it may reduce to a truism. If this construal is feasible, then the internalism-externalism debate will ultimately be a matter of terminology, as far as linguistics is concerned. This may appear a plausible conclusion, in light of the fact that, as pointed out by Bilgrami (1992: 2), Owens (1992: 89) and Rowland et al (2020: Section 2), among others, externalism emerged as an antithesis to the Cartesian idea that "the mental processes had by an individual are always, and completely, determined by what is going on inside that individual's body" (Rowland et al. 2020: Section 2). This Cartesian idea can be pushed so far as to entail what Putnam (1975: 136) called 'methodological solipsism', namely the view that "no psychological state presupposes the existence of the subject's body even: if P is a psychological state, properly so called, then it must be logically possible for a 'disembodied mind' to be in P". The Cartesian assumption conflicts with cognitive linguistics, where physical properties of language users are viewed as essential to the understanding of the nature of language to the extent that cognition is "grounded in perception and bodily experience" (Langacker 2008: 28). Langacker shares with the externalist the anti-Cartesian idea that "[a]n individual's notion of what an expression means develops through communicative interaction" (Langacker 2008: 30).

6. Concluding Remarks

Properly understood, externalism is likely to be better placed than internalism from the very outset, because the former makes merely an existential claim while the latter commits one to a stronger, universal claim (Hunter 2003: 733). What Putnam's and Burge's thought experiments are designed to establish is that there *can* be twins who are molecularly identical but have different concepts and thoughts or, equivalently, that a person's intrinsic properties *can* fail to individuate his concepts and thoughts vis-à-vis his twin's (cf. Brown 2004: ix, 4). In order to refute this view, it must be shown that there can be no two people who are internally identical but have different concepts and thoughts, or, in Segal's (2000: 28) words, that "[a]ny twin of mine in any environment would have a concept with the same extension condition". This contrast is emphasized by Haukioja (2017) and Rowlands et al. (2000):

> […] internalism holds that propositional content is *always* supervenient on intrinsic properties – externalism, as the negation of internalism, holds that the corresponding expressions (i.e. expressions which are phonetically or orthographically identical) used by duplicates may *at least in some* cases differ in *some* of their semantic properties" (Haukioja 2017: 866, emphases in the original).

> […] externalism is the view that an individual's bodily (including neural) events, states, processes, etc. do not always, on their own, determine the mental, states, processes had or undergone by that individual. The qualifier *not always* permits there to be cases – perhaps many – in which mental occurrences are fixed by bodily occurrences. Externalism, generically at least, is merely committed to the claim that this does not *always* happen. The qualifier *on its own* is intended to convey that even if an individual's bodily occurrences do not *completely* determine which mental occurrences

are had or undergone by that individual, the former will at least *partially* determine the latter and, indeed, can play an important role in such determination. (Rowlands et al. 2020: Section 1, emphases in the original)

It follows that, as Hunter (2003: 733) states, "the opponent of Internalism has the prima facie easier case". Suppose that on Twin Earth there is iron pyrites but no gold and that iron pyrites is called 'gold' in Twin English. Suppose also that Oscar has no ability to distinguish gold and iron pyrites, both of which are found on his planet, Earth. Assuming that Oscar and Oscar$_2$ are molecularly identical, this implies that their psychological states as individualistically described are indistinguishable when they talk or think about what they call 'gold'. Still, it will be reasonable to suppose that Oscar's utterance of 'gold', but not Oscar$_2$'s utterance of 'gold', can refer to gold. Since Oscar$_2$ has never had any contact with gold nor with those who have had contact with gold, there is no means for him to think about gold, and his utterance of 'gold' refers exclusively to iron pyrites. The sameness of the twin speakers' internal state does not make the meanings of their 'gold' identical. If Oscar$_2$ were to visit Earth to sell what he called 'gold' as gold, he would be charged with fraud. This single example will suffice to refute the claim that there can be no two people who are internally identical but have different concepts and thoughts. Segal (2000: 27) refers to the internalist view quoted above that "[a]ny twin of mine in any environment would have a concept with the same extension condition" as "[t]he less extreme version" of internalism. The most extreme version of internalism, 'radical internalism', "dispenses with the notions of extension and extension conditions altogether" (Segal 2000: 27)[21]. Faced with the challenge set by Putnam's and Burge's influential thought experiments, the internalist may even have trouble defending a moderate version of his own claim.

References

Abbott, Barbara (1997) "A Note on the Nature of 'Water'", *Mind* 422: 311-319.

Bach, Kent (1987) *Thought and Reference*, Oxford: Clarendon Press.

Barber, Alex (ed.) (2003) *Epistemology of Language*, Oxford: Oxford University Press.

Benveniste, Émile (1956/1966) « La nature des pronoms », in *For Roman Jakobson*, Den Hague: Mouton, reprinted in Benveniste (1966), *Problèmes de linguistique générale I*, Paris: Gallimard, 251-257.

Ben-Yami, Hanoch (2001) "The Semantics of Kind Terms", *Philosophical Studies* 102: 155-184.

Bilgrami, Akeel (1992) *Belief and Meaning*, Cambridge: Basil Blackwell,

Bloomfield, Leonard (1933) *Language*, London: George Allen.

Brown, Jessica (2004) *Anti-Individualism and Knowledge*, Cambridge, MA: MIT Press.

Burge, Tyler (1979/2007) "Individualism and the Mental", *Midwest Studies in Philosophy* 4: 73-121, reprinted in Burge (2007), pp. 100-150.

Burge, Tyler (1982/2007) "Other Bodies", A. Woodfield (ed.), *Thought and Object*, Oxford UP, pp. 97-120,

---

[21] Segal (2000: 157, n. 2) notes that radical internalism is held by Chomsky and "some other prominent linguists". Radical internalism is not accepted by every linguist, of course. Thus, Higginbotham (1998: 160) says that "semantics is a matter of knowledge of conditions on reference".

reprinted in Burge (2007), pp. 82-99.

Burge, Tyler (1988) "Individualism and Self-Knowledge", *Journal of Philosophy* 85: 649-663.

Burge, Tyler (1989/2007) "Wherein Is Language Social?", Alexander L. George (ed.), *Reflections on Chomsky*, Blackwell, pp. 175-191, reprinted in Burge (2007), pp. 275-290.

Burge, Tyler (1993) "Concepts, Definitions, and Meaning", *Metaphilosophy* 24: 309-327, reprinted in Burge (2007), pp. 291-306.

Burge, Tyler (2007) *Foundations of Mind*, Oxford: Oxford University Press.

Chomsky, Noam (1955) "Logical Syntax and Semantics: Their Linguistic Relevance", *Language* 31: 36-45.

De Brabanter, Philippe and Bruno Leclercq (2019) "Proposition d'une enquête empirique sur les intuitions « externalistes » des locuteurs à travers le mode de déférence sémantique", *Travaux du Cercle Belge de Linguistique-Studies van de Belgische Kring voor Linguïstiek* 13: 1-17.
https://difusion.ulb.ac.be/vufind/Record/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/285669/Holdings

Davidson, Donald (1987) "Knowing One's Own Mind", *Proceedings and Addresses of the American Philosophical Association*, 60: 441-458.

Deutsch, Harry (1993) "Semantics for Natural Kind Terms", *Canadian Journal of Philosophy* 23: 389-411.

Donnellan, Keith S. (1983/2012) "Kripke and Putnam on Natural Kind Terms", In Carl Ginet & Sydney Shoemaker (eds.) *Knowledge and Mind*, Oxford: Oxford University Press, 1983, pp. 84-104, reprinted in Joseph Almog and Paolo Leonardi (eds.) Essays on Reference, Language, and Mind, Oxford: Oxford University Press, 2012, pp. 178-203.

Donnellan, Keith (1993) "There Is a Word for That Kind of Thing: An Investigation of Two Thought Experiments", *Philosophical Perspective* 7: 155-171.

Farkas, Katalin (2006) "Semantic Internalism and Externalism", in Ernst Lepore and Barry C. Smith (eds)
The Oxford Handbook of Philosophy of Language, Oxford: Oxford University Press, pp. 323-340.

Fodor, Jerry A. (1975) *The Language of Thought*, New York: Thomas Y. Crowell.

Fodor, Jerry A. (1994) *The Elm and the Expert*, Cambridge, MA: MIT Press.

Frege, Gottlob (1892/1997) "Über Sinn und Bedeutung", *Zeitschrift für Philosophische Kritik*, 100, pp. 25-50. English version in: M. Beaney, ed. (1997). *The Frege Reader*. Oxford: Blackwell, pp. 151-171.

Frege, Gottlob (1918-1919/1956) "Der Gedanke: Eine logische Untersuchung", *Beiträge zur Philosophie des deutschen Idealismus* 1: 58-77, G. Frege (1956) "The Thought: A Logical Inquiry", *Mind* 65: 289-311.

Gasparri, Luca and Diego Marconi (2021) "Word Meaning", *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2021/entries/word-meaning/>.

Grabarczyk, Paweł (2016) "How 'Meaning' Became 'Narrow Content'", *Studies in Logic, Grammar and Rhetoric* 46(59): 155-171.

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

Hacking, Ian (2007b) "Natural Kinds: Rosy Dawn, Scholastic Twilight", *Royal Institute of Philosophy Supplement* 61: 203-239.

Häggqvist, Sören and Wikforss Åsa (2018) "Natural Kinds and Natural Kind Terms", *British Journal for the Philosophy of Science* 69: 911-933.

Häggqvist, Sören and Wikforss Åsa (2020) "The Need for Descriptivism", in Stephen Biggs and Heimir Geirsson (eds.) *The Routledge Handbook of Linguistic Reference*, London: Routledge, pp. 335-344.

Haukioja, Jussi (2017) "Semantic Externalism and Semantic Internalism" In *A Companion to the Philosophy of Language*, Second Edition, ed. B. Hale, C. Wright, and A. Miller, 865-880. Malden, MA: Wiley-Blackwell.

Higginbotham, James (1998) "Conceptual Competence", *Philosophical Issues* 9: 149-162.

Hunter, David (2003) "Gabriel Segal, *A Slim Book About Narrow Content* (MIT Press, 2000), 177 pp.", *Noûs* 37: 724-745

Kaplan, David (1989) "Demonstratives", *Themes from Kaplan*, ed. by J. Almog, J. Perry, and H. Wettstein, 481-563, Oxford: Oxford University Press.

Kripke, Saul A. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.

Lakoff, George (1987) *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, University of Chicago Press.

Langacker, Ronald W. (2008) *Cognitive Grammar: A Basic Introduction*, Oxford University Press, Oxford.

Linsky, Bernard (1977) "Putnam on the Meaning of Natural Kind Terms" *Canadian Journal of Philosophy* 7: 819-828.

Liu, Jeeloo (2002) "Physical Externalism and Social Externalism: Are they Really Compatible?", *Journal of Philosophical Research* 27: 381-404.

McGeer, Victoria (1994) "*Belief and Meaning*. by Akeel Bilgrami", *Journal of Philosophy* 91: 430-439.

McGinn, Colin (1977) "Charity, Interpretation, and Belief", *Journal of Philosophy* 74: 521-535.

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

Needham, Paul (2011) "Microessentialism: What is the Argument?", *Noûs* 45: 1-21.

Noonan, Harold (1984) "Fregean Thoughts", *The Philosophical Quarterly* 34: 205-224.

Nuccetelli, Susana (ed.) (2003) *New Essays on Semantic Externalism and Self-Knowledge*, Cambridge, MA: MIT Press.

Owens, Joseph (1992) "Psychophysical Supervenience: Its Epistemological Foundation", *Synthese* 90: 89-117.

Putnam, Hilary (1962) "The Analytic and the Synthetic", *Minnesota Studies in the Philosophy of Science* 3: 358-397.

Putnam, Hilary (1973) "Meaning and Reference", *Journal of Philosophy* 70: 699-711.

Putnam, Hilary (1975) "The Meaning of Meaning", *Minnesota Studies in the Philosophy of Science* 7: 131-193.

Putnam, Hilary (1990) "Is Water Necessarily H_2O?", In James Conant (ed.) *Realism with a Human Face*, Cambridge, Mass.: Harvard University Press, p. 54-79.

Putnam, Hilary (1992) "Replies", *Philosophical Topics* 20: 347-408.

Quine, W.V.O (1951) "Main Trends in Recent Philosophy: Two Dogmas of Empiricism", *Philosophical Review* 60: 20-43.

Recanati, François (1993) *Direct Reference*, Oxford: Blackwell.

Robertson Ishii, Teresa and Philip Atkins (2020) "Essential vs. Accidental Properties", *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2020/entries/essential-accidental/>.

Rowlands, Mark, Joe Lau, and Max Deutsch (2020) "Externalism About the Mind", *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2020/entries/content-externalism/>.

Sainsbury, Mark and Michael Tye (2012) *Seven Puzzles of Thought and How to Solve them: An Originalist Theory of Concepts*, Oxford: Oxford University Press.

Sakai, Tomohiro (2022a) "Toward an Internalist Construal of Semantic Externalism", Vassiliaudou, Hélène and Marie Lammert (eds.) *A Crosslinguistic Perspective on Clear and Approximate Categorization*, New Castle upon Tyne: Cambridge Scholars Publishing, pp. 23-50.

Sakai, Tomohiro (2022b) "Incomplete Understanding Without Analyticity: Prototype Semantics and Social Externalism", *Tokyo University Linguistic Papers* (*TULIP*) 44, October 2022, pp. e102-e132.

Segal, Gabriel M. A. (2000) *A Slim Book About Narrow Content*, Cambridge MA: MIT Press.

Soames, Scott (2002) *Beyond Rigidity: The Unfinished Semantic Agenda of* Naming and Necessity, Oxford: Oxford University Press.

Stalnaker, Robert (1993) "Twin Earth Revisited", *Proceedings of the Aristotelian Society* 93: 297-311.

Stich, Stephen P. (1978) "Autonomous Psychology and the Belief Desire Thesis", *Monist* 61: 573-591.

Wiggins, David (1995) "Putnam's Doctrine of Natural Kind Words and Frege's Doctrines of Sense, Reference and Extension: Can they Cohere?", in John Biro and Petr Kotatko (eds.), pp. 59-74.

Wikforss, Åsa (2008) "Semantic Externalism and Psychological Externalism", *Philosophy Compass* 3: 158-181.

Wikforss, Åsa (2013) "Bachelors, Energy, Cats and Water: Putnam on Kinds and Kind Terms", *Theoria* 79: 242-261.

Woodfield, Andrew (2000) "Reference and Deference", *Mind & Language* 15: 433-451.

外在主義とは何でないか

酒井 智宏

t-sakai@waseda.jp

キーワード: 外在主義　ステレオタイプ　指標性　本質主義　意味のありか

要旨

　言語学者の多くは外在主義に関心を示さず、外在主義は多くの誤解に晒されてきた。この論文の目的はそうした誤解のいくつかを払拭することにある。第一に、外在主義は言語理解に心的なものが関与しないという立場ではない。パトナムの言うステレオタイプは言語表現の外延を決定するために不十分であるものの、言語使用者の頭の中にある。第二に、外在主義は「水」のような自然種名が「私」のような指標詞の仲間であるとは考えない。「私」の指示対象は文脈に依存するが、「私」の言語的意味は文脈とは独立に定まっている。これに対して、外在主義が正しければ、「水」の言語的意味はこの語が導入された物理的環境に依存する。第三に、外在主義は必ずしも本質主義を含意するわけではない。クリプキは可能世界、固定指示、(形而上学的)必然性といった様相概念を用い、「水は本質的に(= すべての可能世界で、必然的に)H_2O である」と主張したが、パトナムの双子地球の思考実験が成立するためにそうした強い様相的主張は必要ない。第四に、外在主義は意味と思考の内容が物理的に心の外にあるという立場ではない。外在主義は意味と思考の内容が言語使用者/思考主体の外的環境を考慮に入れない限り決定されない(ことがある)という立場である。言語学は外在主義に背を向け、内在主義/個体主義を採用する傾向があるが、これまでのところ外在主義があらゆる点で言語学と無縁であることは示されていない。

<div align="right">（さかい・ともひろ　早稲田大学)</div>